

Université de Montréal.

Décisions et Perspectives :

De la théorie de l'utilité à la philosophie de la volonté.

Par

Olivier Contensou, sous la codirection de François Lepage et Yvon Gauthier.

Département de Philosophie,
Faculté des arts et des sciences.

Thèse présentée à la Faculté des arts et des sciences comme exigence partielle en vue de
l'obtention du grade de docteur en philosophie.

Août 2013.

©, Olivier Contensou 2013.

Thèse soutenue le 07 avril 2014 devant le jury suivant :

Frédéric Bouchard : président-rapporteur.

Alain Voizard : examinateur externe.

Peter Dietsch : examinateur.

François Lepage : directeur.

Yvon Gauthier : codirecteur.

Décisions et Perspectives :

De la théorie de l'utilité à la philosophie de la volonté.

Résumé

Les implications philosophiques de la Théorie de la Perspective de 1979, notamment celles qui concernent l'introduction d'une fonction de valeur sur les résultats et d'un coefficient de pondération sur les probabilités, n'ont à ce jour jamais été explorées. Le but de ce travail est de construire une théorie philosophique de la volonté à partir des résultats de la Théorie de la Perspective. Afin de comprendre comment cette théorie a pu être élaborée il faut étudier la Théorie de l'Utilité Attendue dont elle est l'aboutissement critique majeur, c'est-à-dire les axiomatisations de la décision de Ramsey (1926), von Neumann et Morgenstern (1947), et enfin Savage (1954), qui constituent les fondements de la théorie classique de la décision. C'est entre autres la critique – par l'économie et la psychologie cognitive – du principe d'indépendance, des axiomes d'ordonnancement et de transitivité qui a permis de faire émerger les éléments représentationnels subjectifs à partir desquels la Théorie de la Perspective a pu être élaborée. Ces critiques ont été menées par Allais (1953), Edwards (1954), Ellsberg (1961), et enfin Slovic et Lichtenstein (1968), l'étude de ces articles permet de comprendre comment s'est opéré le passage de la Théorie de l'Utilité Attendue, à la Théorie de la Perspective. À l'issue de ces analyses et de celle de la Théorie de la Perspective est introduite la notion de *Système de Référence Décisionnel*, qui est la généralisation naturelle des concepts de fonction de valeur et de coefficient de pondération issus de la Théorie de la Perspective. Ce système, dont le fonctionnement est parfois heuristique, sert à modéliser la prise de décision dans l'élément de la représentation, il s'articule autour de trois phases : la visée, l'édition et l'évaluation. À partir de cette structure est proposée une nouvelle typologie des décisions et une explication inédite des phénomènes d'akrasie et de procrastination fondée sur les concepts d'aversion au risque et de surévaluation du présent, tous deux issus de la Théorie de la Perspective.

Mots clés : décision, volonté, utilité, Théorie de la Perspective, Système de Référence Décisionnel, akrasie, procrastination.

Abstract

The philosophical implications of Prospect Theory (1979), notably those concerning the introduction of a value function on outcomes as well as a weighting function on probabilities, have not been explored until now. The goal of this thesis is to construct a philosophical theory of the will based on the results of Prospect Theory. In order to understand how Prospect Theory was arrived at, it is necessary to study the Theory of Expected Utility of which it is the critical culmination, in particular the mathematical treatments of decision and utility by Ramsey (1926), von Neumann & Morgenstern (1947), and Savage (1954), which together constitute the foundations of classical decision theory. It was primarily the critique – from economics and cognitive psychology – of the axiom of independence, the ordering axiom and the transitivity axiom that gave rise to the subjective representational elements on the basis of which Prospect Theory was elaborated. These criticisms were made by Allais (1953), Edwards (1954), Ellsberg (1961), and Slovic & Lichtenstein (1968); studying their articles will elucidate how the transition from Expected Utility Theory to Prospect Theory occurred. Following these analyses, the concept of a ***Decisional Frame of Reference*** is introduced, which is the natural generalization of Prospect Theory's concepts of a value function and a weighting function. This system, whose function is sometimes heuristic, serves to model decision-making in the element of representation, and consists of three phases: decisional scope, edition, and evaluation. Based on this structure a philosophical theory is proposed, including a new typology of decisions and an original explanation of the phenomena of weakness of will (akrasia) and of procrastination founded on the concepts of risk aversion and present overweighting, both of which stem from Prospect Theory.

Keywords: decision, will, utility, Prospect Theory, Decisional Frame of Reference , weakness of will (akrasia), procrastination.

Table des matières

Introduction générale	2
1) <u>Motivations philosophiques</u>	2
2) <u>Méthodologie</u>	10
 Chapitre I : La théorie classique de la décision	 16
Introduction	17
 A - Probabilités et décision chez Ramsey	 22
1) <u>Le subjectivisme</u>	22
2) <u>La définition opératoire de la croyance et le concept de pari</u>	25
3) <u>Les propositions éthiquement neutres et l'axiomatique</u>	33
 B - Décision et utilité chez von Neumann et Morgenstern	 50
1) <u>Le jeu comme paradigme décisionnel et le problème de l'utilité</u>	50
2) <u>Une nouvelle conception de l'utilité et de la décision</u>	54
3) <u>L'axiomatisation de l'utilité</u>	59
a – Ordonnancement	64
b – Ordre et combinaison : <i>indépendance et continuité</i>	67
+ <u>Indépendance</u>	67
+ <u>Continuité</u>	70
c – Combinaison et algèbre	73
 C - Décision et incertitude : la théorie de Savage	 77
1) <u>Les définitions préliminaires et le découpage ensembliste</u>	77
2) <u>Les cinq premiers axiomes</u>	84
Axiome 1	85
Axiome 2	87
Axiome 3	91

Axiomes 4 et 5	94
3) <u>Les probabilités subjectives et l'utilité</u>	97
Axiome 6	104
Axiome 7 et le problème de l'utilité	106
Chapitre II : du paradoxe d'Allais à la Théorie de la Perspective	112
Introduction	113
A - Les premières critiques de la TUA	116
1) <u>La valeur psychologique des perspectives</u>	116
<u>et le problème de la pondération</u>	
2) <u>Paradoxe et rationalité</u>	125
B - L'approfondissement de la critique	139
1) <u>La synthèse critique de Ward Edwards</u>	139
2) <u>Le paradoxe d'Ellsberg et le problème du contexte décisionnel</u>	147
3) <u>Le problème du renversement des préférences</u>	155
C - La Théorie de la Perspective	166
1) <u>Les heuristiques et les biais</u>	166
2) <u>La Théorie de la Perspective de Kahneman et Tversky (1979)</u>	178
a – Une critique renouvelée de la TUA	178
b – L'élaboration de la théorie	192
+ <u>La phase d'édition</u>	192
+ <u>Le rôle de la fonction de valeur</u>	198
+ <u>Le coefficient de pondération</u>	201
c – Les effets croisés des deux coefficients	207
et les extensions de la TP	
Chapitre III : Le Système de Référence Décisionnel	214
Introduction	215

A - Rationalité, contexte et environnement décisionnel	220
1) <u>La double limitation de la rationalité</u>	220
2) <u>Du contexte à l'environnement</u>	230
a – Insuffisance du concept de contexte	230
b – Le découpage de l'environnement décisionnel	234
3) <u>La résistance et les orientations de l'environnement</u>	239
 B - De la théorie de la décision à la philosophie de la volonté	 248
1) <u>Décision et représentation</u>	248
a – La déconnexion du décidable et du volontaire	248
b – SRG et représentation	253
2) <u>SRD et fonction de valeur</u>	256
a – Avoir présent et état présent	256
b – L'aversion aux pertes et l'aversion au risque	263
3) <u>La généralisation du phénomène de pondération</u>	270
a – Le travail du SRD	270
b – Mesure probabiliste et évaluations probabilistes	276
 C - Temporalité, édition et évaluation	 282
1) <u>Pondération et perception du temps en contexte décisionnel</u>	282
a – Le caractère essentiellement dynamique de toute décision	282
b – La pondération du temps	288
+ <u>La contrainte de temps</u>	289
+ <u>La procrastination</u>	291
+ <u>Altérité, affectivité et temporalité</u>	295
2) <u>Le système de référence décisionnel et la phase d'édition</u>	297
a – La structure logique des perspectives et des alternatives	297
b – La construction des alternatives	303
+ <u>Le codage affectif des perspectives et des alternatives</u>	305
+ <u>Le rôle des croyances dans l'édition des perspectives</u>	311
3) <u>L'évaluation et la prise de décision</u>	319

a – Délibération, statu quo et akrasie	319
+ <u>Édition et évaluation</u>	319
+ <u>L'élimination des alternatives</u>	322
<u>et les quatre règles de l'évaluation</u>	
+ <u>Évaluation modale et évaluation axiologique</u>	324
+ <u>L'évaluation en termes de gains et de pertes</u>	329
<u>et le problème de l'akrasie.</u>	
b – Les différents types d'évaluation et l'apprentissage	334
+ <u>L'évaluation automatique</u>	334
+ <u>L'évaluation heuristique</u>	335
+ <u>Le SRD et le concept d'apprentissage</u>	338
 Conclusions générales	 344
 Bibliographie	 358

Formalisme utilisé, abréviations, et références.

1) Symboles logiques et mathématiques

- \wedge : conjonction ; $p \wedge q$: p et q .
- \vee : disjonction ; $p \vee q$: p ou q .
- \neg : négation ; $\neg p$: *il n'est pas le cas que p , non p .*
- \supset : implication matérielle ; $p \supset q$: *si p alors q .*
- \equiv : biconditionnel ; $A \equiv B$: *A si et seulement si B .*
- \exists : quantificateur existentiel ; $\exists x$: *il existe un x , quelques x .*
- \forall : quantificateur universel ; $\forall x$: *tous les x .*
- $E!$: quantificateur de l'unicité ; $E!x$: *il existe exactement un x .*
- S : ensemble universel ; ou événement universel chez Savage. *Il a tous les ensembles ou tous les événements comme éléments.*
- \in : appartenance à un ensemble ; $a \in A$: *a est un élément, appartient à A .*
- $0, \emptyset$: événement nul chez Savage, ensemble vide. *Il n'a aucun élément.*
- A^c : complémentaire de A dans S . *Tout ce qui dans S n'est pas A .*
- \subset : inclusion stricte ; $A \subset B$: *A est inclus dans B , certains éléments de B ne sont pas dans A . **Attention** : Savage utilise \subset dans le sens de \subseteq .*
- \subseteq : inclusion faible ; $A \subseteq B$: *A est inclus dans B , tous les éléments de A sont dans B . **Attention** : Savage utilise \subseteq dans le sens de \subset .*
- \cap : intersection de deux ensembles ; $A \cap B$: *intersection de A et de B , les éléments de S qui sont communs à A et B .*
- \cup : union de deux ensembles ; $A \cup B$: *A union B , éléments de S qui sont des éléments de A ou de B ou des deux.*
- \bigcap : intersection généralisée ; $\bigcap_i A_i$: *intersection de tous les A_i .*
- \bigcup : union généralisée ; $\bigcup_i A_i$: *union de tous les A_i .*
- \succ : relation de préférence forte ; $a \succ b$: *a est strictement préféré à b .*

- $>$: strictement supérieur à. **Attention** : ce symbole est utilisé également dans les axiomatiques de Ramsey, von Neumann/Morgenstern, et Savage pour exprimer la préférence stricte, auquel cas il doit être lu comme « strictement préféré à ».
- \leq : inférieur ou égal. **Attention** : Ce symbole sert aussi à exprimer la préférence faible dans l'axiomatique de Savage ; $A \leq B$: B est faiblement préféré à A , A n'est pas faiblement préféré à B .
- \sim : relation d'indifférence ; $A \sim B$: A et B sont indifférents.
- \leq^* : probabilité subjective faible ; $A \leq^* B$: A est faiblement plus probable que B au niveau subjectif.
- \sim^* : indifférence quant aux probabilités subjectives ; $A \sim^* B$: A n'est pas plus probable que B au niveau subjectif.
- \doteq : indifférence conditionnelle ; $f \doteq g | B$: l'acte f est indifférent à l'acte g étant donné l'événement B . Autrement dit l'événement B est nul.
- $|$: probabilité conditionnelle ; $P(p|q)$: la probabilité que p étant donné q .
- \sum : somme ; $\sum_{j=1}^n x_j$: la somme des x indice j , quand j va de 1 à n .
- \rightarrow : application de l'ensemble U dans l'ensemble des réels de l'axiomatique von Neumann/Morgenstern.

2) Abréviations.

Comme il y a dans ce travail des noms d'auteurs ou de théories qui reviennent souvent, nous avons choisi d'utiliser un système d'abréviations afin d'alléger le texte de certaines répétitions. Les abréviations sont les suivantes :

- TP : Théorie de la Perspective.
- TUA : Théorie de l'Utilité Attendue.
- TJ : Théorie des Jeux.
- TCP : Théorie Cumulative de la Perspective.
- MUA : Maximisation de l'Utilité Attendue.

- RP : Renversement des Préférences.
- PCS : Principe de la Chose Sûre.
- PEN : Proposition Éthiquement Neutre.
- PRI : Principe de Raison Insuffisante.
- HB : Heuristiques et Biais.
- SRD : Système de Référence Décisionnel.
- SRG : Système de Référence Général.
- N/M : von Neumann/Morgenstern.
- K/T : Kahneman/Tversky.
- L/S : Lichtenstein/Slovic.

3) Références.

Dans le même esprit que pour les abréviations et afin d'alléger autant que possible le texte, nous avons choisi un système de références pour les articles et les ouvrages collectifs qui fait l'économie des prénoms des auteurs. Par exemple : Fishburn/Wakker (1995) sert d'abréviation à : Fishburn, P., Wakker, P. (1995) : « The Invention of the Independence Condition for Preferences ». *Management Science*. Vol. 41, No. 7, p 1130-1143. Le titre complet se trouve en bibliographie. La combinaison de ce système avec les abréviations introduites plus haut peut produire des références du type : K/T (1979) qui est l'abréviation de : Kahneman, D., Tversky, A. (1979) : « Prospect Theory. An Analysis of Decision under Risk ». *Econometrica*. Vol. 47, No. 2, p 263-291. Reprint in Kahneman, D., Tversky, A. (eds.) (2000), p 17-43.

Ce travail est dédié à Maude, qui illumine ma vie et me rend chaque jour meilleur,

À René Delbos, qui m'a appris à lire, à écrire, et un peu à penser,

À Stéphane et Jérôme Bost, qui savent bien pourquoi,

*À Jeffrey John Hanneman (1964-2013) qui m'a inspiré et « discrètement » accompagné ces
vingt cinq dernières années.*

Remerciements

Je remercie François Lepage et Yvon Gauthier qui ont assumé la codirection de cette thèse et m'ont prodigué de précieux conseils sans ménager aucunement leur temps ni leur engagement. Je remercie particulièrement Mr Lepage pour sa générosité concernant la recherche et l'octroi de bourses d'études.

Je remercie également le groupe « existence et connaissance des objets mathématiques » (FQRSC. Directeur : Mathieu Marion) pour l'attribution généreuse de deux bourses, ainsi que la Faculté des Études Supérieures pour l'octroi d'une bourse de fin de rédaction de dix mois, et l'octroi de deux bourses d'excellence. Mes remerciements vont aussi au département de philosophie de l'Université de Montréal pour son soutien moral et financier de 2009 à 2013.

Je tiens à remercier particulièrement Messieurs David Piché et Daniel Dumouchel pour leur excellente gestion des Études Supérieures, ainsi que pour l'octroi de deux bourses de recherches. Mes remerciements vont aussi particulièrement aux professeurs Alain Petit, Emmanuel Cattin, et Élisabeth Schwartz pour avoir initié et consolidé en moi la volonté de savoir. Qu'ils trouvent ici témoignage de ma profonde reconnaissance pour leurs extraordinaires enseignements.

Je tiens à remercier Jean-Pierre Marquis, qui est pour moi une source d'inspiration philosophique et pédagogique, ainsi que Mathieu Marion pour son fort soutien, ses conseils et ses encouragements, ainsi que son point de vue sur la philosophie.

Enfin, je tiens à remercier ma compagne Maude qui m'a soutenu durant toute cette rédaction, avec un Amour, une présence et une constance qui ont contribué à la qualité de mon travail, tout comme la pertinence et l'intelligence de ses commentaires et remarques. Mes remerciements vont aussi à Caroline Contensou, Antoinette et Mireille Ibanès pour leur soutien inconditionnel durant ces vingt années de philosophie, et pour toute la joie qu'elles mettent dans ma vie depuis toujours. Pour terminer, je tiens à remercier spécialement Louis-Hervé Barichard, Christian Lachance, Gilles R. Cadrin, Germain Marsan et Adam Westra, qui de par leurs amitiés, leurs styles, et leurs apports philosophiques, dans et par la pratique de la dialectique, contribuent à cultiver en moi la volonté de savoir.

« Il faut poser des actes d'une si complète audace, que même ceux qui les réprimeront devront admettre qu'un pouce de délivrance a été conquis pour tous. »

Claude Gauvreau

« ... La liberté naîtra, corps adulte accouché par l'infiniment petit piétiné. Le grossier a des membres, le grotesque a des bras ; réel sans lourdeur appréciable, réel inaperçu que l'on néglige. Ce qui n'a pas été vu, ce qui ne sera pas vu facilement grossira sous forme de soleil. Les armées du désir purifiant, panorama intangible d'une précurSION intuitive. Fédralbor turiptulif, corne de muse agrippée au cosmos. Libualdivane, drétlôdô cammuef ; l'élixir des archanges toisonne au fond des crêtes. Liberté-rides aqueuses... »

Claude Gauvreau.

Introduction générale

1) Motivations philosophiques.

Le but de ce travail est de construire une théorie de la volonté à partir des acquis de la Théorie de la Perspective (TP) élaborée par Kahneman et Tversky en 1979¹. Théorie qui est une généralisation critique de la Théorie de l'Utilité Attendue (TUA)². L'idée directrice de cette thèse sera donc de s'interroger sur la possibilité de passer d'une théorie de la décision à caractère formel vers une théorie de la volonté d'une texture plus philosophique et conceptuelle. Il s'agira de *dépasser de façon conceptuelle les limitations de la TUA* en remplaçant le décidable sous l'égide unificatrice du volontaire, puisque la théorie de la décision avait séparé, pour les raisons du formalisme, décision et volonté. Or nous pensons que cette séparation coupe la décision de toute sa densité subjective et représentationnelle, puisque nous définissons la volonté comme *la décision dans l'élément de la représentation*³. La réintroduction du concept de volonté sera rendue nécessaire à partir du moment où nous aurons démontré que les décisions sont, non seulement accompagnées par des représentations d'ordres divers, mais aussi construites à partir de représentations subjectives (affects, valeurs, croyances, connaissances, souvenirs, etc.). Le concept central qui nous permettra de définir précisément l'acte de la volonté, entendue comme un ensemble de processus, est le *Système de Référence Décisionnel* (SRD), introduit et développé au chapitre III.

Comme nous allons nous en rendre compte au cours de nos analyses du chapitre I, la TUA, représentée ici par les travaux de Ramsey (1926), von Neumann/Morgenstern (N/M) (1944), et Savage (1954), est insuffisante pour expliquer de façon exhaustive des décisions humaines. Bien que très abouties du point de vue formel⁴, ces théories ne parviennent pas à capturer ce qui fait l'essence même d'une décision, à savoir l'intervention systématique de

¹ Kahneman, D., Tversky, A. (1979) : « Prospect Theory : An Analysis of Decision under Risk ». *Econometrica*, 47:2, 263-91.

² Théorie dont les deux principaux réquisits sont que les options ont une certaine valeur quantifiable et arrivent avec des probabilités objectives ou subjectives. Le choix s'opère donc en multipliant la valeur de l'option par sa probabilité, c'est donc l'alternative qui aura l'utilité attendue la plus haute qui sera choisie. Il s'agit donc de maximiser l'utilité attendue. La TP, bien que modifiant sensiblement l'aspect mathématique de la TUA, notamment par l'intervention de deux coefficients, restera très proche de cette dernière.

³ Nous nous occuperons ici seulement de la décision humaine, nous laissons de côté le problème de la décision animale.

⁴ Voir Simon (1955) et Fishburn (1970).

représentations subjectives. En effet le formalisme est obligé de laisser de côté un bon nombre d'éléments, par exemple l'humeur ou les émotions du sujet, dont on sait qu'elles interviennent de façon première dans les décisions¹.

Les axiomatiques que nous allons étudier dans le chapitre I, sauf celle de N/M – qui est selon Savage seulement descriptive² – ont des prétentions *normatives* élevées. C'est-à-dire qu'elles entendent établir comment les sujets *devraient* décider réellement. Or cette prétention à la normativité est la plupart du temps³ exorbitante. Ce qui nous amène à nous poser cette question : en connexion avec les problèmes de normativité, quelles sont les raisons qui peuvent nous pousser à mettre en doute la pertinence philosophique de la TUA pour expliquer et prédire les décisions humaines telles qu'elles sont prises dans la réalité ? Qu'est-ce qui fait la faiblesse conceptuelle de la TUA lorsqu'il s'agit de rendre compte par exemple des décisions subjectives habituelles ou quotidiennes ?

Il y a tout d'abord le présupposé selon lequel les sujets, du fait qu'ils doivent maximiser ou qu'ils maximisent, sont des *calculateurs*. Outre un aspect machinique, peut-être un peu irréaliste, la possibilité de calculer et d'appliquer une axiomatique peut être sérieusement remise en question pour au moins deux raisons : tous les sujets ne sont pas capables au même degré d'accéder aux axiomatiques et de les intégrer : elles demandent une éducation mathématique et un outillage formel d'un niveau assez élevé⁴. Le fait que tout le monde puisse développer la motivation nécessaire et suffisante pour accéder à ce niveau est en lui-même douteux. Pourtant cela n'empêche pas les sujets de gérer leurs situations décisionnelles usuelles de façon pertinente⁵. Deuxième raison, même en levant l'hypothèque de la possibilité pour les sujets d'accéder aux axiomatiques, il y a des situations dans lesquelles, ni les probabilités, ni la valeur des options n'interviennent, des situations qui sont

¹ Voir entre autres Zajonc (1980) ; de Sousa (1987) ; et Damasio (1995). Ce qui ne veut pas dire que la TUA n'a subi aucun affinement ni aucune modification pour intégrer ces éléments, comme le montre par exemple l'entreprise de Hermalin/Isen (2006).

² Voir Savage (1972), p 97.

³ Limitation importante, car il existe des cas, comme les « Dutch book » où l'aspect normatif – en l'occurrence ici du calcul des probabilités – peut éviter au sujet de perdre tout ce qu'il possède.

⁴ C'est ce qui ressort des analyses sur la rationalité limitée de Simon (1955) et Gigerenzer (2011), mais à un autre point de vue : les sujets n'ont pas besoin d'autant de rationalité pour décider. Les standards sont placés trop haut par les défenseurs de la TUA. Par ailleurs, Ellsberg (1961), montrera que même lorsque l'on signale aux sujets que leurs préférences violent certains axiomes – comme le Principe de la Chose Sûre par exemple – ils ne ressentent pas le besoin d'amender leurs choix, ils ne trouvent pas que leurs préférences soient irrationnelles.

⁵ Il reste néanmoins que l'on peut se demander dans quelle mesure la théorie de la décision peut-elle améliorer la prise de décision ? Des auteurs comme Davidson ou Lewis se sont confrontés à cette question, nous y reviendrons à la fin de notre analyse de l'axiomatique N/M. Pour un point de vue pertinent en français sur la question, voir Mongin (2011).

donc par définition incalculables, même de façon métaphorique. Dit autrement, dans l'hypothèse où les sujets auraient intégré les axiomes de la TUA et seraient alors devenus des similis robots décisionnels, il existe un ensemble de situations – que nous allons analyser – pour lesquelles le formalisme reste lettre morte. Notamment les situations habituelles et inhabituelles non critiques.

Comme nous le notions plus haut, la TUA « désubjectivise » toute décision, et elle opère une déconnexion entre le décidable et le volontaire justement parce qu'elle fait l'économie d'éléments représentationnels centraux dans le domaine du choix rationnel. Or, nous entendons montrer que c'est le lien entre le sujet et ses représentations – subsumées sous le concept de Système de Référence Décisionnel – qui permet d'expliquer l'ensemble, de saisir « le grain fin »¹, des processus impliqués dans le fonctionnement de la volonté. C'est pourquoi aussi nous allons postuler un Système de Référence Général (SRG) qui inclut en lui l'ensemble des éléments cognitifs définissant le sujet comme ce sujet-ci : les affects, les valeurs, les croyances, les connaissances, les souvenirs. Le système de référence décisionnel est un sous-système du SRG².

Cependant est-ce à dire qu'à l'issue de nos analyses tous les concepts et axiomes de la TUA auront été évacués comme non pertinents ? Nullement, certains d'entre eux, notamment l'*axiome de continuité*, garderont une portée descriptive et normative dans le cadre de notre théorie, par exemple pour formaliser le rapport que les sujets entretiennent avec le risque de façon quotidienne.

Dans le chapitre deux, nous allons étudier les critiques de la TUA, notamment au travers des paradoxes d'Allais et d'Ellsberg³, mais aussi au travers de la problématique du renversement des préférences mise au jour par Lichtenstein et Slovic dans les années 70⁴. Le but de cette étude sera double. Il s'agira tout d'abord de montrer comment les expériences ont mis en faillite les axiomes de la TUA en faisant émerger des *éléments subjectifs* intervenant de façon systématique dans le choix rationnel, éléments que le formalisme avait échoué à intégrer. Ces éléments ce sont par exemple le rapport des sujets aux probabilités, au risque et à la certitude. C'est ce que mettrons en avant les critiques d'Edwards (1954) et d'Allais (1953) :

¹ Pour reprendre une idée de Pascal Engel (1997), p 339, qui nous beaucoup inspirée et que nous développerons.

² Pour plus de clarté, nous renvoyons le lecteur à notre schéma, p 219.

³ Voir Allais (1952), (1953) ; Ellsberg (1961).

⁴ Voir Slovic/Lichtenstein (1968) ; Lichtenstein/Slovic (1971) et (1973).

les probabilités sont pondérées et déformées du fait de l'application de coefficients déterminés par la façon dont les sujets se *représentent* les probabilités. De leurs côtés, Ellsberg (1961) et Slovic/Lichtenstein (1968), mettront en faillite le concept classique de contexte décisionnel, ainsi que les axiomes d'ordonnancement et de transitivité. *Tous ces éléments critiques seront à verser au registre de l'intervention de la subjectivité dans les décisions*, subjectivité que la TUA avait précisément bannie. La plupart de ces critiques, notamment celles d'Allais et d'Ellsberg, vont surtout avoir pour cible l'*axiome d'indépendance*, qui est le cœur de la TUA et sur lequel est fondée l'hypothèse de maximisation¹.

À notre sens, le but d'une théorie de la décision *réaliste*, doit être de pouvoir intégrer les éléments qui sont à la source des paradoxes, comme l'aversion au risque et l'attrait des sujets pour les alternatives certaines, le renversement des préférences en fonction des gains et des pertes, ou de la tâche à effectuer². C'est précisément le but que se propose la Théorie de la Perspective (TP) de Kahneman et Tversky (1979), et cela nous conduit à la deuxième raison qui a motivé l'étude des critiques de la TUA dans notre deuxième chapitre : ces critiques préparent l'avènement de la TP, à partir de laquelle nous avons construit notre propre théorie de la volonté et de la décision.

L'étude de la TP constitue l'aboutissement de notre partie analytique, et le point d'Archimède de tout le travail, puisque c'est à partir des acquis de cette théorie que va être construit le Système de Référence Décisionnel. Cette théorie, qui dans sa forme reste très proche de la TUA représente l'achèvement à la fois critique et conservateur de cette dernière. Il s'agira donc ici de sonder et d'explorer les *implications philosophiques* de la Théorie de la Perspective, ainsi que d'élaborer à partir de cette dernière une structure permettant d'expliquer l'ensemble des décisions humaines, de la dimension semi-automatique, aux décisions ultimes.

La théorie de Kahneman et Tversky, étayée sur un ensemble de protocoles expérimentaux, arrive à capturer cinq éléments essentiels à partir desquels nous allons reconstruire toute la dynamique décisionnelle. Ces éléments sont : l'*édition* des perspectives, leur *évaluation* à partir de l'avoir présent du sujet, l'introduction de la *fonction de valeur* qui formalise le rapport du sujet à son avoir présent et la dissymétrie gains/pertes, l'introduction d'un *coefficient de pondération* sur les probabilités qui formalise la façon dont le sujet se

¹ Fishburn (1970), p 108, et la démonstration de Picavet (1996), p 160-163.

² Voir Lichtenstein/Slovic (1971), in Lichtenstein/Slovic (Eds.) (2006), p 63.

représente ces dernières, et enfin le phénomène de l'*aversion au risque* qui se manifeste dans le fait que *les pertes pèsent toujours plus lourd que les gains*. Ces cinq éléments sont robustes du point de vue expérimental, certains d'entre eux avaient déjà fait l'objet d'investigations, mais il revient à la TP de les subsumer sous une seule structure mathématique.

C'est à partir d'une analyse et d'une réorganisation conceptuelle de ces cinq éléments que nous allons élaborer notre propre modèle, qui aura une portée plus générale. Plus particulièrement, c'est à partir de la fonction de valeur – parce qu'elle représente le point de vue à partir duquel le sujet évalue les perspectives – que vont être pensés le SRG et le SRD. Une question se pose néanmoins : puisque la TP parvient à intégrer les éléments subjectifs sources de paradoxes qu'avait mis en avant la critique, pourquoi une nouvelle théorie d'ordre plus philosophique et d'une portée plus générale est-elle nécessaire ? La réponse à cette question est assez simple : c'est justement parce que la TP est *insuffisante du point de vue conceptuel et représentationnel* qu'elle doit être prolongée et généralisée par une théorie du Système de Référence Décisionnel. Deux raisons principales peuvent expliquer cet état de fait.

Tout d'abord la TP ne permet pas de construire le concept de *visée*, ou *environnement décisionnel* dans lequel s'inscrit la représentation du but à atteindre. Dans cette théorie, les alternatives sont proposées aux sujets par le biais de questionnaires. Il y a donc un certain artificialisme dans la façon dont les perspectives sont présentées aux sujets¹, perspectives qui vont être ici reconstruites via la phase d'édition, mais qui ne seront pas originellement élaborées par le sujet. Or c'est un problème majeur puisqu'un nombre important de décisions repose sur la construction de la visée par le sujet, construction qui déclenche le processus d'édition des alternatives². Une théorie réaliste de la décision doit pouvoir expliquer comment les sujets construisent et se représentent leurs buts, en fonction de leur état intérieur et de l'information extérieure.

Deuxième raison qui justifie le dépassement de la TP : cette théorie se maintient dans un contexte probabiliste classique, le contexte de risque, où l'utilité des alternatives est quantifiable et où ces dernières arrivent avec des probabilités objectives. Or, il appert que les décisions qui sont accompagnées d'une mesure probabiliste objective sont assez rares dans

¹ Artificialisme qu'avait anticipé Savage (1972), notamment en se demandant dans quelle mesure les réponses aux questionnaires seraient un reflet réel de la façon dont les sujets se comporteraient dans la réalité.

² Voir notre schéma, p 219.

l'existence, même si elles sont possibles. Nous devons nous doter d'un *modèle qui puisse permettre de penser les décisions non probabilistes*, qui ne peuvent être soumises au calcul. Il faut donc aussi établir une typologie des décisions, non plus en fonction de leur contexte – qui reflète originellement une information probabiliste – mais en fonction de l'*aspect de la situation*, ou de l'état du monde auquel elles sont liées : *habituel non-critique, habituel critique, inhabituel non-critique, inhabituel critique, exceptionnel*. Remarquons aussi que c'est la qualité de l'état du monde en question qui va justifier l'intervention d'une mesure probabiliste ou pas. Par exemple dans le cadre d'une situation habituelle non-critique une telle mesure ne sera pas nécessaire, alors que dans une situation de type inhabituelle critique oui.

Enfin, il est remarquable que bien que restant dans une optique économique, la TP, par l'établissement de la fonction de valeur, soit parvenue à thématiser que c'est l'avoir présent du sujet qui doit être pris en compte et non plus *la seule considération de l'état futur* comme dans la TUA. C'est un progrès décisif en ce sens qu'il nous met sur la voie de l'intégration d'un plus grand nombre de facteurs essentiels pour toute prise de décision, à savoir les affects, les émotions et l'humeur du décideur, mais aussi tous les autres éléments qui font de lui un sujet totalement singulier et individualisé. Cependant pour pouvoir intégrer ces éléments il va falloir sortir du contexte économique et se situer dans une structure conceptuelle élargie, dont le socle sera la fonction de valeur, ainsi que les phases d'édition et d'évaluation.

Cette structure conceptuelle sera exposée au chapitre III, il s'agit du Système de Référence Décisionnel (SRD), qui est lui-même subsumé sous un système plus large : le Système de Référence Général (SRG)¹. L'idée d'un *système de référence* incluant en lui l'ensemble des processus intervenant dans les décisions est issue de la relativité restreinte, et en particulier de l'idée de « point de vue » sur la réalité associée au concept de système de référence physique². Ce qu'il s'agit de faire ressortir ici, c'est la singularité irréductible, qui détermine le sujet comme ce sujet-ci, sur laquelle fait fond toute décision. Cette singularité est due à la *configuration* du Système de Référence Général dans lequel sont inclus les éléments représentationnels et cognitifs qui de par leurs relations, leur organisation et leur hiérarchisation, font du sujet un sujet unique. Ces éléments sont les croyances, les affects, les valeurs, les connaissances et les souvenirs. Tous peuvent avoir une influence plus ou moins

¹ Pour plus de clarté, nous renvoyons à notre schéma, p 219.

² Par exemple dans l'article « sur l'électrodynamique des corps en mouvement », paragraphe 3.

importante dans les processus décisionnels, tous ces éléments sont interconnectés. Ce sont justement ces éléments que bannissait la TUA des processus décisionnels et que ne prend que partiellement en compte la TP.

La question qui constitue le fil directeur de notre troisième chapitre est la suivante : une fois admis que la TUA et la TP ne peuvent formaliser qu'un petit nombre de décisions, à savoir les décisions à caractère probabiliste, comment le SRD doit-il être constitué pour pouvoir expliquer toutes les autres formes de décisions humaines ? Quels sont le rôle et la portée explicative du SRD ? En fin de compte quel programme de recherches devons-nous nous fixer pour arriver à avoir une vision exhaustive du travail de la volonté ?

Le Système de Référence Décisionnel est la généralisation philosophique *naturelle* de la TP, en particulier de la fonction de valeur. Lorsque nous parlons de généralisation philosophique nous entendons ici appliquer les concepts introduits dans la TP dans un cadre plus large où le formalisme ne fonctionne plus pour la simple raison que le nombre de facteurs pertinents à intégrer serait trop important, et que ce sont des facteurs qui se laissent mal formaliser. Le but est de pouvoir comprendre comment la plupart de nos décisions, sinon toutes, s'articulent autour des trois opérations que sont *l'établissement de la visée*, *l'édition* et *l'évaluation* ; mais aussi nous avons pour vocation d'expliquer comment le SRD se *stabilise* autour de certaines heuristiques et habitudes décisionnelles, phénomène qui peut avoir différents effets tels que la surpondération du statu quo ou des résistances plus pathologiques comme *l'inertie décisionnelle* qui peut conduire à la *procrastination* ou à *l'akrasie*.

Il est apparu dans le cours de nos recherches que le SRD est un système *conservateur*, notamment dans son fonctionnement habituel, c'est-à-dire un système qui bien que pouvant évoluer, s'auto-évaluer et s'autoréguler¹, a pour vocation de stabiliser le régime décisionnel des sujets autour de certains pôles habituels qu'il faudra analyser. Ce fait s'explique en partie parce que c'est un système écologique et économique, au sens où Simon (1955) et Gigerenzer/Brighton (2009) l'entendent, à savoir que cette économie de moyens est le reflet d'une rationalité limitée de façon endogène et exogène. Mais c'est aussi le reflet d'une certaine tendance – d'ordre plus général – qu'à l'existence à se stabiliser autour de pôles habituels, cette stabilisation a un impact immédiat sur le SRG et par là même sur le SRD.

¹ Il a donc certaines des caractéristiques d'un système métacognitif, voir Proust (2007), cette piste qui ne sera pas approfondie dans ce travail mériterait un traitement exhaustif.

Il est très important aussi que le SRD puisse refléter de façon exacte comment les affects, les émotions et l'humeur du décideur interviennent dans les processus décisionnels. Cette intégration devra prendre en compte les avancées récentes en philosophie et en psychologie qui montrent que les affects n'ont pas seulement un rôle négatif à jouer dans la construction des décisions¹. Mais plus généralement il faudra repenser le rapport des représentations subjectives aux décisions, et par exemple réévaluer le rôle des croyances dans la constitution du choix rationnel. En fait c'est l'effet possible de l'ensemble des représentations cognitives dans la sphère du décidable qui doit être réévalué.

C'est pour arriver à ce but que nous avons fait l'hypothèse d'un Système de Référence Général qui inclut en lui, en plus du SRD, tous les éléments représentatifs susceptibles d'accompagner aussi bien l'établissement de la visée, que les phases subséquentes d'édition et d'évaluation. Mais pour parvenir à ce but il faudra donc aussi analyser comment les éléments représentationnels que sont les croyances, les valeurs, les connaissances, les souvenirs et les affects interagissent, c'est-à-dire comment, de par leurs liaisons et relations ils forment un *système subjectif totalement unique*.

Nous suivons ici l'intuition de Pascal Engel lorsqu'il émet l'hypothèse que pour comprendre l'action il faut arriver à analyser de façon aussi fine que possible comment le sujet lui-même s'investit dans ses représentations. Nous allons appliquer cette recommandation méthodologique dans le domaine de la décision parce que représentation et décision sont des concepts liés de façon analytique. Faire l'économie de l'aspect représentationnel du décidable c'est renoncer à comprendre la dynamique décisionnelle. Nous aurons l'occasion de nous rendre compte en conclusion que c'est précisément dans cet investissement et dans l'assentiment que le sujet peut accorder ou refuser à certaines représentations cognitives que se situe la liberté².

Le SRD est un système qui doit pouvoir expliquer tous les types de situations décisionnelles qu'un sujet peut rencontrer, et quelles stratégies sont mises en places pour répondre aux dites situations. Nous faisons, à la suite de Gigerenzer³ mais aussi de Kahneman

¹ Nous pensons par exemple à de Sousa (1987) ; Damasio (1995) ; Tappolet (1995), (2003) et (2010).

² Nous nous situons donc ici dans une optique très proche de celle du Stoïcisme ancien, qu'il faudra pour le coup réactualiser. Le problème de la liberté et de l'éthique ne sera abordé qu'en conclusion.

³ Voir par exemple Gigerenzer (1996), et Gigerenzer/Brighton (2009).

et Tversky¹, l'hypothèse d'un fonctionnement heuristique du SRD. Il conviendra donc de mettre au jour ce que nous allons appeler les « heuristiques d'édition » et les « heuristiques d'évaluation », en particulier dans le fonctionnement quotidien du SRD, puisque ces heuristiques sont le reflet du fonctionnement écologique et économique de ce système.

Enfin, une explication originale de la procrastination et de l'akrasie sera proposée, reliée au concept d'aversion au risque et au fait que les pertes pèsent plus lourds que les gains, tous deux directement issus de la Théorie de la Perspective. L'idée est assez simple : le SRD est un système conservateur, mais qui a la capacité d'apprendre et d'évoluer pour faire face à la dynamique normale du SRG ainsi qu'aux situations diverses qui peuvent se présenter dans l'existence ; cependant des cas peuvent se présenter où la surpondération, la surévaluation du présent et des habitudes peuvent produire une *inertie pathologique*. Nous pensons que la procrastination et l'akrasie sont deux façons qu'a cette inertie de se manifester, l'une lors de la phase d'édition, l'autre lors de la phase d'évaluation.

2) Méthodologie.

Il s'agit d'un travail qui se situe au carrefour de deux styles philosophiques coordonnés par un objectif commun : construire un modèle philosophique réaliste et falsifiable² de la volonté. Les deux méthodes que nous avons utilisées ici peuvent être, avec toutes les réserves qui s'imposent, qualifiées d'« *analytique* » et de « *conceptuelle* ». Clarifions immédiatement ces notions. Lorsque nous parlons de méthode analytique nous entendons ici analyse dans un sens assez littéral. Il s'agit en effet dans les deux premiers chapitres d'analyser de façon la plus précise possible la littérature scientifique pertinente qui constitue le fondement de la théorie moderne de la décision. Ces théories étant très formelles, il a fallu analyser l'aspect logico-mathématique de chacune d'elles, et faire parfois un peu de reconstruction logique lorsque le formalisme présenté par les auteurs pouvait prêter à confusion ou n'était pas aussi clair qu'il aurait dû l'être³. Nous avons tenté d'expliciter autant que possible l'ensemble des

¹ Voir Tversky/Kahneman (1974).

² Nous pensons qu'il est en effet possible de tester certaines parties du modèle de façon empirique.

³ C'est le cas par exemple de l'axiome d'Archimède chez Ramsey, de l'axiome d'indépendance chez von Neumann/Morgenstern, ou de l'explicitation des prérequis mathématiques à la construction d'une théorie de la mesure que nous présentons dans notre partie sur Savage.

théories que nous avons analysées en les contextualisant et en les comparant entre elles, ainsi qu'en faisant, autant qu'il était possible et dans les limites du raisonnable, une revue de littérature qui est à la fois la plus complète et la plus à jour. Ceci explique en partie que l'appareil de notes soit aussi fourni surtout dans les deux premiers chapitres.

Par ailleurs nous avons tenu aussi à garder le formalisme original de chacune des théories, parfois en le modernisant un peu, mais nous tenions à préserver cet aspect qui nous semble important, car il permet d'individuer les théories en question¹ et de percevoir le « style » déployé dans les analyses au sens que Granger donne à ce terme², de percevoir dans quelles mesures ces œuvres, bien que ressortissant de l'esprit scientifique portent en elles comme la signature de leurs auteurs³.

Les analyses du formalisme sur lequel se fonde la théorie classique de la décision étaient nécessaires, car ce n'est qu'à partir de la critique de ce formalisme que vont émerger progressivement les éléments sur lesquels nos propres hypothèses seront construites. Tout le deuxième chapitre est consacré à cette problématique, c'est-à-dire à la mise au jour – par les contestations empiriques – d'éléments subjectifs que la TUA ne parvenait pas à formaliser et qui constituent pourtant des éléments fondamentaux de la décision humaine comme le rapport des sujets au risque et aux probabilités.

L'introduction des concepts de fonction de valeur, coefficient de pondération, phase d'édition et d'évaluation, aversion au risque, n'a été rendue possible que par les analyses qui furent menées par les économistes d'une part, notamment Allais, et les psychologues d'autre part, notamment Edwards. L'intégration de ces concepts dans une théorie synthétique, la Théorie de la Perspective, se fera dans un premier temps sur le mode du formalisme. Formalisme qui lui aussi sera analysé de près à la fin du chapitre deux, qui constitue la dernière analyse formelle de la thèse, puisque c'est à partir de ce moment que s'opère le passage à la deuxième méthode que nous utilisons et que nous avons baptisée « conceptuelle ».

¹ La plupart des livres qui portent sur la théorie de la décision normalisent le formalisme, pour les raisons de l'exposition, et parce que cela permet aussi de présenter les symboles une fois pour toutes, voir par exemple, Fishburn (1970), Kreps (1988), Picavet (1996), Peterson (2009), Wakker (2010), etc.

² Voir Granger (1988), par exemple p 8, qui définit le style comme « *modalité d'intégration de l'individuel dans un processus concret qui est travail* ». Ce livre sera d'ailleurs l'occasion pour Granger d'analyser le style de Savage dans la construction de sa théorie des probabilités subjectives.

³ Voir Ricoeur, 1986, p 120, qui cite d'ailleurs le texte de Granger que nous mentionnons ci-dessus.

Que faut-il entendre ici par « méthode conceptuelle » ? Méthode qu'inaugure le troisième chapitre. Il va s'agir de se placer en continuité stricte avec les analyses qui ont été menées dans les deux premiers chapitres, mais en donnant aux concepts convoqués, notamment ceux de la Théorie de la Perspective une densité rigoureusement philosophique. Il va s'agir par exemple de répondre à des interrogations du type : de quelle façon faut-il reconstruire et penser la notion de fonction de valeur – directement issue de la TP – pour qu'elle devienne un concept philosophique ayant une portée suffisante pour penser la structure des décisions humaines ? Autrement dit, par quels moyens pouvons-nous passer d'une notion mathématique à sa généralisation philosophique, sans que celle-ci ne semble trop artificielle ou trop peu solide ? Pour ce faire il va falloir passer à un autre niveau d'analyse, qui tout en se détachant de la littéralité des textes, va se fixer un autre type de limite, qui est aussi une question : quels sont les éléments représentationnels qui sont convoqués dans les décisions, et comment interagissent-ils à la fois entre eux et avec le Système de Référence Décisionnel. Il va donc s'agir de *reconstruire le volontaire par l'analyse des relations entre décision et représentation*. C'est l'interrogation qui constitue le fil directeur du troisième chapitre.

C'est une interrogation qui est à la fois *conceptuelle*, dans le sens où elle cherche à élaborer, à fonder, un concept nouveau – le SRD – à partir d'une structure mathématique, mais l'interrogation est aussi *exploratoire*, car elle doit parvenir à constituer les composants du SRD à partir des éléments de la représentation. Le cheminement est le suivant : à partir des acquis de la TP nous posons l'existence d'un Système de Référence Décisionnel comme l'extension philosophique naturelle de cette théorie, nous posons donc un concept, mais nous devons ensuite – pour donner au concept ainsi posé toute l'extension nécessaire – explorer et déployer les parties qui le constituent comme tel, à savoir ici les concepts de *visée*, d'*édition* et d'*évaluation*, mais aussi explorer les éléments desquels il dépend quant à son fonctionnement immédiat, à savoir le SRG et les éléments représentationnels qu'il inclut, ainsi que le rapport du SRG au flux informatif extérieur. Pour mener cette enquête à bien, il faut donc par exemple analyser des concepts comme *la résistance*, puisque s'il n'y avait pas de résistance il n'y aurait pas de décision, décider c'est toujours vaincre une forme de résistance, aussi minime soit-elle. Il faut aussi analyser le rapport que le SRD entretient avec le temps, ce que ne font ni la TUA

ni la TP¹, et c'est une carence majeure. Nous aurons l'occasion de nous rendre compte que le rapport du décidable au temps et au concept de délai est majeur.

La façon dont la question est posée, ainsi que les notions qui sont convoquées dans ce chapitre nous obligent donc à opter pour une méthode conceptuelle définitionnelle et exploratoire dont la caractéristique principale pourrait être qu'elle ne tourne le dos à aucune théorie, ni à aucune discipline, du moment que cette théorie est pertinente pour la constitution et l'exploration du domaine considéré. C'est ce qui explique notre utilisation de certains acquis issus de la phénoménologie pour construire notre concept d'environnement décisionnel, et les références que nous pouvons faire – tout au long de ce travail – aux traditions continentales et analytiques.

Dans un domaine aussi complexe que la philosophie de la décision et de la volonté, tous les points de vue sont unilatéraux, et toutes les disciplines insuffisantes pour parvenir à une vision exhaustive et réaliste du phénomène. Peut-être que la « vérité » du décider humain se situe au confluent des points de vue auxquels nous conduisent la logique, les mathématiques, l'économie, la psychologie, et la phénoménologie. Et ce point de vue ne peut être que celui d'une philosophie qui se sera départie de toute unilatéralité, de toute partialité, qui aura renoncé au territorialisme ou à un régionalisme par trop exacerbé². C'est exactement dans cet esprit qu'a été entrepris et mené ce travail, d'où en fin de compte aucun point de vue n'est exclu pour des raisons autres que de pertinence.

Pour pouvoir prétendre à une réelle expertise – qui ne doit pas être confondue avec la fermeture d'un territoire épistémique sur lui-même, bien au contraire –, il est nécessaire de se situer dans une perspective où les disciplines communiquent entre elles, une perspective *pluri disciplinaire*³ pour des problématiques – telles que celle du choix rationnel – qui sont *transdisciplinaires*. Le problème de la décision humaine traverse toutes les disciplines susmentionnées, dont on pourrait dire qu'elles forment le socle, ou le fondement d'une

¹ Il y a eu cependant des tentatives pour intégrer l'aspect dynamique et temporel à la théorie de la décision, comme en témoignent les travaux d'Edwards (1962), et de McClennen (1990).

² D'une façon plus générale, il faudrait d'ailleurs se demander dans quelle mesure un régionalisme philosophique trop exacerbé – pour des raisons la plupart du temps de doctrine, d'école ou de tradition – et dont la spécialisation serait devenue l'excuse, n'obscurcirait pas plus qu'il n'éclairerait notre compréhension de la réalité, qui se donne à nous de façons multiples impliquant par là même des grilles de lectures diverses mais complémentaires. Ce qui ne veut pas dire que les analyses de détail doivent être rejetées, bien au contraire. Mais il faut arriver à articuler dans un même regard le particulier, le précis, les analyses fines avec une structure conceptuelle plus générale ayant une portée plus importante et plus réaliste.

³ Elle pourrait se situer dans l'optique que développe Gilles R. Cadrin dans son livre de 2011 : *L'être humain face à lui-même. L'animal autoréférent*.

approche interdisciplinaire de la décision. Mais les choses pourraient encore se continuer, se complexifier et se nuancer, notamment en allant dans des directions théoriques où le décider est central comme la politique ou le droit, mais aussi en se demandant si les animaux sont dotés des éléments cognitifs et représentatifs suffisants pour que leur soient attribuées des décisions¹.

Nous allons faire trois dernières remarques avant de clore cette introduction. Dans un premier temps nous voudrions mettre en relief la structure interne des deux premiers chapitres, qui est analytique dans son déroulement, mais synthétique lorsque l'on considère les chapitres comme des totalités, puisque les théories qui clôturent ces deux chapitres sont à chaque fois la synthèse des théories qui les précèdent dans l'exposition : Savage opère une synthèse entre Ramsey et von Neumann/Morgenstern, la Théorie de la Perspective propose une synthèse de l'ensemble des travaux qui vont de Allais à Lichtenstein et Slovic. Mais la Théorie de la Perspective opère en fait une double synthèse, puisqu'en restant dans le paradigme de la TUA elle synthétise aussi les théories de Ramsey, von Neumann Morgenstern et Savage. Ce qui fait que l'analyse de la TP représente en fait la synthèse de l'ensemble des deux premiers chapitres, elle constitue comme nous le disions, à la fois, pour des raisons philosophiques et méthodologiques, le centre de gravité de l'ensemble du travail.

Deuxième point. Tout au long de ce travail il aura fallu se situer entre deux options théoriques, que nous considérons comme des écueils partiels : la thèse ricoeurienne de la « monstruosité » du concept de fait psychique², autrement dit de l'impossibilité absolue de réifier quoi que ce soit du décidable, à cet égard les réussites localisées, mais bien réelles, de la théorie de la décision infirment le point de vue ricoeurien. Mais, il fallait aussi éviter de tomber dans l'écueil inverse qui eut consisté à tenter de formaliser l'ensemble du vouloir, à cet égard, l'échec de la théorie de la décision à mathématiser l'ensemble du décidable donnerait raison à Ricoeur de façon partielle. Nous avons choisi de nous situer entre ces deux thèses unilatérales, mais complémentaires. Pour la simple raison qu'à notre sens c'est la seule position qui permette de se tenir le plus proche possible de l'expérience et de la façon dont s'opèrent les décisions humaine dans leurs réalités et leurs singularités. Comme nous le verrons dans notre premier et notre deuxième chapitre, il y a une partie du décidable qui est

¹ Voir à ce sujet, entre autres, Proust (2010).

² Voir Ricoeur (1950), p 12.

formalisable, mais elle ne représente qu'une partie très circonscrite des décisions que les sujets prennent dans leurs existences.

Dernier point, tout au long du travail, mais surtout dans les deux premiers chapitres nous avons opté pour un appareil de notes assez lourd, qui répond en fait à différentes exigences que nous nous sommes fixés. La première fut d'être autant exhaustif que possible en ce qui concerne les références à produire, notamment lorsque nous avons exposés la TUA et ses critiques. Certaines notes bien que très longues étaient nécessaires dans le souci de faire des revues de littérature aussi précises et à jour que possibles. Les notes peuvent aussi servir à expliciter des points qui bien que non essentiels pour l'argumentation générale sont importants à soulever, et à envisager. Les notes du troisième chapitre gardent leurs fonctions pour les références, cependant elles ont surtout une visée explicative, elles peuvent être considérées comme un ensemble de scolies qui permet d'éclaircir des points un peu complexes de l'exploration conceptuelle, ou de proposer des pistes et des problématiques qui ne peuvent pas être développées dans le corps du texte.

Chapitre I

La théorie classique de la décision :

Ramsey, von Neumann/Morgenstern, Savage.

Introduction.

Il est coutumier de faire remonter l'origine de la théorie de la décision à certains éléments de l'histoire de la culture européenne, notamment mathématiques et philosophiques. On confond souvent la naissance d'une protothéorie de la décision avec la naissance du calcul des probabilités. Or il n'en est rien : le concept de probabilité n'apparaît ni dans la solution de Pascal au « problème des partis »¹, ni dans celle de Fermat, ni dans leur correspondance de 1654. Si le « problème des partis » qui consiste à déterminer la juste part qui revient à deux joueurs à l'issue d'une partie interrompue peut être considéré comme un problème de théorie de la décision ou de théorie des jeux, il n'est en aucun cas constitutif d'une théorie des probabilités au sens contemporain du terme. Le problème traité par Pascal et Fermat constitue surtout une théorie de l'équité, qui déborde nécessairement le domaine strictement ludique² et se prolonge dans des considérations commerciales et juridiques³. Cependant dans le type de problème soulevé, le hasard n'est pas absent, ce qui vaudra à la discipline le nom de « géométrie du hasard »⁴. Il s'agit donc d'une formalisation séminale de la décision en contexte d'incertitude.

Dans le domaine de la philosophie, c'est le Pari de Pascal et le quatrième chapitre de la *logique ou l'art de penser* de 1662 qui sont considérés comme des textes fondateurs de la théorie de la décision⁵. Le Pari est le premier texte où est introduit le concept de dominance et où sont associés les concepts d'utilité et d'état du monde. Pascal donne ici une mesure quantitative de la conséquence des actes qui dans son esprit doit avoir une portée normative : parce que l'acte de croire à une utilité infinie par rapport à tous les autres et n'est dominé par aucune alternative, même dans les cas où Dieu n'existe pas, la conclusion qui doit s'imposer à

¹ Sur le « problème des partis », la littérature est plus qu'abondante, on peut se reporter notamment à Coumet (1970), p 574-580 ; Hacking (2007), p 57-62.

² Domaine dont l'investigation remonte assez loin dans le temps comme le montre David (1955), en analysant le *De Vetula*, qui est un des premiers textes (Xe siècle) où il est question des combinaisons possibles de deux dés. Des éléments de réflexion sur les jeux de hasard se trouvent aussi dans le *Liber de ludo aleae* de Cardan (1564) et dans le court texte de Galilée : *Sopra le scoperte dei dadi* (1620).

³ Voir Coumet (1970) et Daston (1989), Gigerenzer et Al. (1989), p 19-26. L'émergence de la prise en compte du hasard dans les affaires humaines est aussi liée au contexte socioculturel de l'époque : début de la navigation au long cours, des assurances, etc.

⁴ Le nom vient du *Traité du triangle arithmétique* de Pascal (1654)

⁵ Voir Hacking (2007), p 63-84 ; et Hacking (2004), p 127-135. Il faut noter que dans le premier ouvrage Hacking traite les idées d'Arnaud et Nicole à la suite de Pascal, car la logique en question a une ascendance pascalienne importante.

l'athée libertin est alors de croire¹. La croyance en Dieu constitue ainsi l'acte qui a la plus grande utilité attendue, il est donc ici explicitement question de maximisation. Toutefois, les liens entre le probable et la décision ne sont pas thématiques par Pascal, il faudra attendre *la logique ou l'art de penser* pour que le lien soit établi : le jugement sur les « éléments futurs » doit être conditionné par « la probabilité de l'événement »². À cet égard, nous pouvons affirmer que les Messieurs de Port-Royal disposaient d'une définition du probable philosophiquement plus élaborée que celle de Pascal et plus proche de ce que deviendra le concept en théorie moderne de la décision.

Le mathématicien hollandais Christiaan Huygens continuera et approfondira les réflexions de Pascal et Fermat sur le problème des partis dans un court ouvrage³. Il introduit le concept de pari pour évaluer l'équité des partis, mais le concept central de son ouvrage est celui d'« expectation », dont la valeur est équivalente à la multiplication des prix par le nombre de cas dans lesquels chaque gain est susceptible d'être obtenu. Cette idée était déjà présente de façon implicite dans les travaux de Pascal et de Fermat du fait de la position du problème : à quoi les joueurs ont-ils le droit de s'*attendre* en toute équité ? Cependant il faudra attendre l'*Ars conjectandi*⁴ de Jacob Bernoulli pour obtenir à la fois une synthèse de tout ce qui s'est fait dans la discipline depuis Cardan et une avancée majeure avec notamment l'introduction de la loi faible des grands nombres.

L'ouvrage se divise en quatre parties, et c'est la quatrième partie – que l'auteur n'a pu achever – qui est intéressante pour la théorie de la décision, elle peut être assimilée à une première interprétation philosophique des probabilités en rapport avec les prises de décision, dans cette partie, outre la démonstration de la loi faible des grands nombres, l'auteur établit le lien de façon tout à fait explicite cette fois entre probabilités et choix rationnel, comme il le souligne lui-même :

¹ Comme le remarque Hacking (2004), p 125-126, le pari est une très bonne illustration du concept de dominance et de sa portée normative.

² Arnauld et Nicole, *La logique ou l'art de penser*, p 428-429. Le texte est suffisamment intéressant du point de vue philosophique pour que nous le citions en entier : « Pour juger de ce que l'on doit faire pour obtenir un bien, ou pour éviter un mal, il ne faut pas seulement considérer le bien et le mal en soi, mais aussi la *probabilité* qu'il arrive ou n'arrive pas. (...) La crainte d'un mal doit être proportionnée non seulement à la grandeur du mal, mais aussi à la probabilité de l'événement ». Il s'agit là d'une utilisation très voisine du concept de probabilité de celle que vont en faire les théoriciens modernes de la décision, d'autre part, le point de vue exposé est très proche de ceux que nous allons étudier dans la théorie classique de la décision. On trouve ici aussi en germe l'idée que les probabilités des éléments négatifs sont surévaluées par rapport à celle des éléments positifs, or comme nous le verrons c'est un point central de la TP.

³ *De ratiociniis in ludoe alae*, (1657)

⁴ Publié en 1713 à titre posthume.

« Conjecturer quelque chose, est mesurer sa probabilité ; et pour cette raison l'*Art de conjecturer* ou *stochastique*, est défini par nous comme l'art de mesurer exactement les probabilités des choses, afin que nous puissions toujours choisir dans nos actions, ce qui est préférable ou plus habile ; la sagesse du Philosophe et la prudence du Politique sont d'ailleurs dirigées en ce sens »¹.

Cet extrait anticipe beaucoup les idées que nous allons croiser dans ce chapitre, notamment le lien fort, mais problématique² qui unira chez Ramsey, N/M et Savage, les préférences avec les probabilités. L'extrait anticipe aussi des notions qui vont se retrouver dans la mathématique sociale de Condorcet³ et dans celle de Arrow⁴ : l'application du calcul aux affaires politiques⁵. Bernoulli est aussi le premier à proposer de relier l'augmentation de la « certitude morale » avec le nombre des observations, il propose par là des éléments de logique inductive.

Ce qu'il faut noter aussi, c'est que dans cette période fondatrice et préparatoire, les interprétations objectives et subjectives des probabilités sont fondues dans un même objet⁶, par exemple, dans l'*Ars Conjectandi* Bernoulli utilise les deux points de vue qui bien que n'étant pas présentés comme équivalents, ne sont pas présentés non plus comme antagonistes, et d'une certaine façon ce phénomène durera au moins jusqu'à Laplace. Le concept de probabilité désigne alors aussi bien la mesure de la régularité de certains phénomènes physiques, que le degré de croyance qu'un sujet peut attribuer à un fait, ou que la mesure de l'ignorance.

Le concept d'expectation que Huygens avait mis en place à la suite des travaux de Pascal représente ce qu'un sujet peut *rationnellement* attendre d'un jeu ou d'une perspective. En 1713, Daniel Bernoulli, dans son article « Specimen theoriae novae de mensura sortis » va remettre en doute le fait que l'expectation puisse représenter de façon réaliste la mesure du

¹ *La science de la conjecture*. Traduction française, p 213.

² Notamment du fait de l'introduction de l'axiome d'indépendance dans l'axiomatique N/M.

³ *Tableau Général de la Science*, (1847-1849).

⁴ *Social choice and individual value*, (1951).

⁵ Poisson tentera de son côté d'appliquer les probabilités, en particulier la loi des grands nombres à la composition des tribunaux et des jurys populaire dans le but de produire plus d'équité, mais ces vues seront sévèrement critiquées, ce qui mettra d'ailleurs un frein aux tentatives d'application des probabilités en matière juridique et sociale.

⁶ Voir Hacking (2007), p 11-17. Comme le remarque avec pertinence cet auteur : « It is notable that the probability that emerged so suddenly is Janus-faced. On the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background », p 12.

risque des individus¹. L'article en question contient deux résultats importants et liés entre eux : le phénomène de l'utilité marginale décroissante², et l'introduction du concept d'utilité attendue à la place de celui d'expectation. Le « paradoxe de Saint-Pétersbourg »³ permet de démontrer le décalage entre l'expectation objective d'un jeu et son utilité attendue, utilité (« emolumentum ») qui doit être ici entendue au sens de l'usage qu'un individu peut faire d'une somme d'argent par exemple et non comme une simple mesure de sa quantité. Il s'agit donc de montrer que dans la réalité ce n'est pas l'expectation mathématique que le sujet considère comme essentielle, mais l'espérance morale. Bernoulli énonce le problème ainsi :

« Pierre jette en l'air une pièce et refait de même jusqu'à ce que, à la chute de la pièce, face apparaisse pour la première fois : si ceci se produit au premier jet, il doit donner à Paul un ducat, si c'est au second, deux ducats ; au troisième, quatre ; au quatrième, huit et ainsi de suite en doublant à chaque jet le nombre de ducats. On demande quelle est l'espérance de Paul »⁴.

Dans le cadre de ce problème, l'expectation mathématique de Paul est infinie puisque l'espérance E de gain est :

$$E = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2^2} + 4 \cdot \frac{1}{2^3} + \dots + 2^{n-1} \cdot \frac{1}{2^n} + \dots$$

Si nous respectons la logique de l'expectation mathématique, le prix pour jouer à ce jeu devrait être infini ; or très peu de sujets seraient prêts à aller au-delà de 20 ducats pour se lancer dans une partie. Ce comportement qui semble irrationnel si nous nous en tenons à l'expectation est simplement dû à un phénomène qui aura une très grande importance dans notre travail : *l'aversion au risque*. Mais cela ne signifie en aucun cas qu'avec Bernoulli nous sortons de la maximisation, au contraire, et c'est en ce sens que les théories que nous allons étudier peuvent être qualifiées de « bernoulliennes »⁵, seulement la fonction qu'il s'agira de maximiser ne sera plus la simple expectation mathématique, mais l'*utilité attendue*. Or, dans le

¹ On trouve une intéressante analyse de l'article dans Savage (1954), p 91-95, dont nous nous inspirons ici partiellement.

² Voir Bernoulli (1738), p 26.

³ Puisque l'article a été publié par l'Académie des sciences de Saint-Pétersbourg.

⁴ Bernoulli (1738), p 31.

⁵ Voir Allais (1953), p 514-521.

même article, Bernoulli montre que cette utilité est relative à l'avoir du sujet, et que par exemple un ducat pour un riche n'aura pas la même valeur que la même somme pour un mendiant. C'est un fait important, car il s'agit de ramener l'usage et l'utilité à l'avoir présent des sujets, et dans ce cas-là, nous avons une préfiguration de ce que sera la fonction de valeur en PT, qui permet justement de formaliser comment une perspective est évaluée par rapport à l'avoir présent du sujet. Ce concept, de « l'avoir présent » aura une grande importance dans notre troisième chapitre puisque c'est à partir de lui que nous tenterons de définir l'état présent du SRG et du SRD.

Dernier point de cette introduction : en 1763, Bayes va proposer son fameux théorème sur les probabilités inverses qui permet d'estimer les probabilités de la cause d'un événement à partir d'une information incomplète :

« étant donné le nombre de fois qu'un événement inconnu s'est réalisé ou a fait défaut, on demande la chance que la probabilité de sa réalisation lors d'une seule épreuve soit comprise entre deux degrés que l'on puisse assigner »¹.

La question que pose Bayes ici sera reprise par Ramsey et par Savage en ce qui concerne la mesure de l'utilité ou celui des probabilités subjectives. Nous étudierons plus loin la façon dont Savage construit la mesure de probabilité subjective d'un événement de référence en restant parfaitement fidèle à la méthode originale de Bayes qui consiste à établir des partitions aussi fines que possible. Il reviendra donc à Ramsey et à Savage de construire une théorie qui combine l'idée de maximisation de l'utilité attendue avec l'introduction d'une mesure de probabilité subjective, ce qui constitue un raffinement important de la théorie originale. Les réflexions sur l'utilité cesseront à partir de Laplace, qui sera le dernier à réfléchir sur le concept d'espérance morale. C'est Ramsey qui reprendra le cours de cette réflexion en 1926.

¹ Bayes (1763), p 27. Pour une analyse complète de l'article de Bayes, voir Dale (1999).

A - Probabilités et décision chez Ramsey.

1) Le subjectivisme.

L'interprétation subjective des probabilités¹ a connu un essor majeur dans la première moitié du XXe siècle grâce aux travaux de F.P Ramsey (1926), B. de Finetti (1937) et L. J. Savage (1954). L'idée que les probabilités puissent être corrélées avec le degré de croyance n'est pas nouvelle² puisqu'on peut la faire remonter au moins à *La logique ou l'art de penser*, Bernoulli, Bayes et Laplace³.

L'innovation majeure dans le domaine des probabilités subjectives au XXe siècle consiste dans la définition croisée de celles-ci avec une théorie axiomatique de la décision et de l'utilité⁴. La combinaison d'une interprétation subjective et d'une axiomatisation du choix rationnel élaborée sous l'égide du principe de l'espérance mathématique⁵ constitue le fondement de la théorie bayésienne de la décision. L'idée centrale du bayésianisme⁶ est de modéliser la prise de décision des sujets en partant des désirs et des croyances – qui arrivent par degrés – et en allant vers les préférences⁷ qui vont être formalisées pour aboutir à une représentation probabiliste et ensuite vers l'établissement d'une fonction d'utilité que le sujet est censé maximiser. Il s'agit donc de passer du qualitatif au quantitatif avec l'appui d'un certain formalisme mathématique, ce qui ne va pas de soi⁸.

¹ Les travaux sur l'interprétation des probabilités au XXe siècle sont plus qu'abondants. Dans un premier temps on peut se reporter à Gillies (2000), p 50-87 ; Hacking (2004), p 141-249 et (2006) p 11-17 pour l'émergence des points de vue fréquentistes et subjectivistes ; enfin Kyburg/Smokler (1980) pour une présentation des textes fondateurs ; ainsi que Gärdenfors/Sahlin (1988).

² C'est un problème qui va nous intéresser tout au long de notre travail, et qui peut être formulé sous la forme d'une question : les probabilités sont-elles un outil fiable pour mesurer les degrés de croyances ? À ce sujet voir Ellsberg (1961) ; Hacking (1967).

³ Voir, Gigerenzer et Al. (1989) ; Hacking (2004), (2006) et (2009).

⁴ C'est le cas pour Ramsey (1926), p 172 ; et Savage (1954) p 68-104. Le dernier étant d'ailleurs inspiré par le premier. En ce qui concerne de Finetti, les choses sont plus nuancées comme l'atteste de Finetti (1957), p 7, qui hésitait à suivre Savage sur la piste de l'utilité. On peut cependant construire un théorème de la représentation à partir de la structure mathématique élaborée par de Finetti ; voir à ce sujet Gilboa (2009), p 89 et suivantes ; Gillies (2000), p 55-65 ; enfin Jeffrey (1983), p 199 et de façon plus générale (2002), p 85-96.

⁵ Ce principe donne lieu, à l'intérieur du bayésianisme, au phénomène de maximisation de l'utilité attendue (MUA). Ce principe, utilisé par Ramsey et bon nombre de théoriciens de la décision, sera fortement critiqué par Allais (1953), Edwards (1954), nous y reviendrons longuement. Sur la MUA on peut se reporter à Gärdenfors/Sahlin (1988), p 1 et suivantes ; Binmore (2009), p 116-136 ; Hacking (2004), p 91-137, Peterson (2009), p 91-117 ; et enfin Baron (2009), p 311-339.

⁶ Sur le bayésianisme nous renvoyons aux ouvrages de Oaksford/Chater (2006), et Chater/Oaksford (Eds.) (2008)

⁷ Il y a donc une ascendance humienne assez forte et assez constante dans ce domaine d'investigation, voir Picavet (1996), p 52-95, et p 207. Cette ascendance sera explicitement assumée par Savage et implicitement par Ramsey...

⁸ Il s'agit en fait dans les trois théories que nous allons étudier d'établir une véritable théorie de la mesure. Celle de Ramsey se rapproche et anticipe un peu la tentative carnapienne dans le domaine des mesures physiques. Voir à ce sujet Carnap (1966) et

Le premier auteur qui retiendra notre attention dans le cadre de notre analyse des formalisations du choix rationnel et de l'utilité est Frank Plumpton Ramsey. Cet auteur n'a pas eu un impact immédiat dans le domaine de la théorie de la décision, en tout cas pas avant les années 50 et sa redécouverte par Savage, puis dans les années 60-70 avec sa promotion par des auteurs comme Richard Jeffrey¹. Nous allons nous concentrer sur les trois premières parties de l'article « Vérité et probabilité » (« Truth and probability ») écrit en 1926 et publié à titre posthume en 1931².

Avant de commencer l'analyse du texte nous voudrions faire, en guise d'introduction, deux remarques courtes sur Bruno de Finetti qui, presque au même moment que Ramsey et de façon indépendante, proposa lui aussi une théorie des probabilités subjectives³. Nous voudrions tout d'abord souligner que même si les deux auteurs peuvent être considérés comme les fondateurs « simultanés » du mouvement subjectiviste au XXe siècle, cette fondation partagée s'appuie néanmoins sur certaines différences qu'il faut souligner.

Mais commençons par un rapprochement. Ramsey et de Finetti s'accordent pour *faire des probabilités la mesure du degré de croyance des sujets en une proposition, une option ou un événement*. C'est ce que Ramsey nomme au début de son article « la logique de la croyance partielle »⁴. Partielle en ce sens qu'il s'agit d'une logique « des arguments non concluants »⁵, c'est-à-dire d'une logique inductive⁶. Dans cette logique il ne saurait être question de valeurs de vérités, mais de degrés de croyance compris entre 0 et 1. Il y a pour Ramsey, de façon

sur la théorie de la mesure en général : Krantz/Luce/Suppes/Tversky (2006). Cette opération : rendre la mesure de l'utilité possible à partir d'une détermination des préférences sera différente dans les trois axiomatiques que nous allons analyser, à savoir celles de Ramsey (1926), von Neumann/Morgenstern (1947), Savage (1954). Il faut remarquer que von Neumann/Morgenstern ne rentrent pas à strictement parler dans le cadre du paradigme bayésien du fait qu'ils utilisent l'interprétation fréquentiste des probabilités. Voir, von Neumann/Morgenstern (1947), p 19.

¹ Jeffrey (1965/1983). La bibliographie sur Ramsey est moyennement abondante et les sources les plus pertinentes sur cet auteur sont, à l'heure actuelle, la thèse de John M. Vickers (1962) et le livre de Niels E. Sahlin (1990), auxquels on peut ajouter l'article de Bradley (2007).

² L'article couvre une trentaine de pages et est divisé en cinq parties d'inégales longueurs. Les « œuvres complètes » de Ramsey ont été publiées par R. Braithwaite dans un recueil intitulé *The foundations of mathematics*. Dans ce recueil, beaucoup de textes restent à l'état d'ébauches. Dans notre analyse nous utiliserons la traduction française d'Alain Voizard en nous référant à l'original anglais. L'article sera cité : Ramsey (1926), dans la pagination de l'édition des œuvres de Ramsey chez Vrin (2003).

³ Voir, de Finetti (1937), il s'agit d'une série de conférences devenues un article fondateur. Voir aussi Gillies (2000), p 50 et suivantes, l'analyse de cet auteur est intéressante puisqu'il propose d'emblée une comparaison avec Ramsey. Des renseignements précieux se trouvent aussi dans Binmore (2009), p 89 et suivantes. Cependant les articles fondateurs en termes de comparaison entre Ramsey et de Finetti sont ceux de M. C. Galavotti (1989) et (1991).

⁴ Ramsey (1926), p 153.

⁵ Ibid.

⁶ À ce sujet voir Skyrms (1986) ; Baird (1992) ; et Hacking (2004). Du point de vue de la pédagogie Hacking est indépassable ici, cependant les exemples totalement développés qu'utilise Baird, p 46-79, notamment dans le domaine médical, sont excellents.

immédiate, un lien entre ce type de logique nos croyances et nos actions. De son côté, de Finetti expose sensiblement la même thèse dans son article de 1937 : « Le meilleur exemple est celui d'un championnat où le spectateur attribue à chacune des équipes une probabilité de vaincre plus ou moins grande d'après son propre jugement »¹. D'autre part, dans un texte de 1989, de Finetti attaquera avec véhémence le point de vue objectiviste² et c'est ce que Ramsey fera aussi – plus sobrement certes – dans la deuxième partie de son article afin de montrer que la théorie du choix rationnel ne peut être fondée sur autre chose qu'une interprétation subjective des probabilités. Sa cible sera la théorie de Keynes.

Cependant, de Finetti va pousser sa critique de l'interprétation objectiviste beaucoup plus loin que ne le fera Ramsey, car, comme il l'assène dans l'introduction de son ouvrage de 1970 *Theory of probability* : « Probability does not exist »³.

Ce que veut dire ici l'auteur c'est qu'il n'existe pas de phénomènes tels que des fréquences objectives qui seraient indépendantes du sujet. L'objectivation de fréquences est impossible dans la nature ou dans la science : tout peut être ramené et doit être ramené à un degré de croyance⁴. À la différence de Ramsey, de Finetti soutient un subjectivisme fort qu'il considère lui-même « comme la solution extrême du côté du subjectivisme »⁵. Le point de vue de notre auteur est un peu plus nuancé, car il ménage une place, certes avec quelques réserves, à la possibilité d'appliquer l'interprétation fréquentiste – probablement héritée de Boltzmann – en physique « où la probabilité (...) est en fait la même que la fréquence »⁶.

Pour conclure cette brève comparaison, il faut mentionner le fait que les deux auteurs font du concept de pari l'instrument de mesure et d'évaluation des degrés de croyances. Cependant là aussi leurs points de vue s'avèrent légèrement différents : alors que de Finetti propose une théorie des paris sur une base monétaire⁷ concernant de petits montants par rapport à l'avoir total des sujets, Ramsey préfère, en raison de l'ambiguïté du rapport des

¹ De Finetti (1937), p 8.

² De Finetti, (1989), p 173.

³ De Finetti (1970), Vol I, Introduction.

⁴ Cette thèse, forte, ne va pas sans poser de problèmes, notamment épistémologiques : s'il n'existe pas de fréquences objectives comment les lois peuvent-elles émerger? Par ailleurs qu'advient-il de la loi des grands nombres? De Finetti parviendra à résoudre ce problème grâce au concept novateur d'« échangeabilité » qui permet de réinjecter le concept d'objectivité dans les probabilités subjectives. Voir, de Finetti (1937), chapitres III, IV et V ; ainsi que Galavotti (1989) ; Gillies (2000), p 69-77 ; Jeffrey (2004), p 86-88 ; et Binmore (2009), p 42-43.

⁵ De Finetti (1937), p 3.

⁶ Ramsey, op. cit., p 155.

⁷ Voir de Finetti (1937), p 6 ; et Gillies (2000), p 55.

sujets à l'argent et de son utilité marginale décroissante, des paris en termes de « biens » et de « maux »¹.

Ces différences sont importantes, car elles nous permettent de voir qu'à l'intérieur de l'interprétation subjectiviste différentes options théoriques demeurent possibles² qui bien que partant du même principe et des mêmes axiomes³ proposent des représentations probabilistes des croyances assez dissemblables ; avec des enjeux là aussi différents.

2) La définition opératoire de la croyance et le concept de pari.

L'article publié par Ramsey en 1926 occupe une place centrale dans le domaine de la fondation du bayésianisme et de l'axiomatisation de la décision⁴. L'originalité de Ramsey se situe dans quatre dimensions :

- La *réactualisation* du concept de pari comme base de la mesure du degré de croyance et une définition opératoire du concept de croyance⁵.

- Introduction du concept de *proposition éthiquement neutre* (PEN) comme point d'indifférence dans l'échelle des préférences du sujet. Comme nous le verrons, le concept de PEN ne va pas de soi.

- Introduction d'une axiomatique tout à fait originale ayant pour fonction de formaliser l'espace des préférences du sujet entre des options ou des mondes possibles⁶. Cette axiomatique permet de dériver une fonction d'utilité et de calibrer les degrés de croyance de façon probabiliste. Ce formalisme sera analysé de près, car nous allons y trouver, de façon séminale, des éléments importants que nous retrouverons chez von Neumann/Morgenstern et Savage.

- Nous trouvons aussi dans cet article une justification de l'induction qui quoiqu'étant dans l'esprit de Mills est originale, notamment car elle s'inscrit en faux contre les vues du

¹ Ramsey, op. cit., p 175. Il est bien évident que ces concepts devront être clarifiés.

² Savage (1954) et Jeffrey (1965/1983) représentent encore deux autres alternatives du côté du subjectivisme.

³ Axiomes qui sont sensiblement équivalents à ceux que proposera Kolmogorov en 1933. Concernant la différence entre les axiomatiques de Kolmogorov et de Finetti, voir Gillies (2000), p 65 et suivantes ; ainsi que Galavotti (1989).

⁴ Voir, Gärdenfors/Sahlin, p 1 et suivantes.

⁵ Nous parlons bien ici de *réactualisation* car en plus du précédent pascalien où le pari concerne la croyance en Dieu, il y a un précédent kantien important qui cette fois relie le pari à la propension que nous aurions à agir sur la base de nos croyances, nous sommes donc ici beaucoup plus proches de Ramsey. Voir, Kant (1787), p 854 ; Picavet, p 176-180.

⁶ Voir Ramsey, op. cit., p 170.

Tractatus de Wittgenstein assez répandues à ce moment- là¹. Il s'agit pour Ramsey de promouvoir une « logique humaine » ou « logique de la vérité » qui doit pouvoir nous guider dans nos actions et dans l'obtention de nos connaissances.

C'est par une discussion critique sur les interprétations fréquentistes et logicistes² que s'ouvre l'article de Ramsey. Les concessions faites à la théorie fréquentiste sont rapidement présentées par l'auteur³ qui passe ensuite à une critique plus approfondie des idées de John Maynard Keynes dans son ouvrage *A treatise on probability* (1921).

De la même façon que la logique de la croyance partielle ne peut être représentée de façon exhaustive en termes de fréquences, cette logique ne peut non plus être représentée par le rapport, pour le moins problématique, qu'entretiennent entre elles les propositions d'une inférence. Le degré de croyance, et par là même le degré de probabilité, ne peut être exprimé par le rapport entre deux propositions prises comme prémisse et comme conclusion. Or, c'est justement le point de vue soutenu par Keynes, que Ramsey résume ainsi : « Si la relation est de degré α de la croyance totale en la vérité de la prémisse, nous devrions, si nous étions rationnels, passer à une croyance de degré α en la vérité de la conclusion »⁴. Pour notre auteur, il est impossible de faire du rapport entre les deux propositions la définition de la probabilité ainsi que l'entend Keynes : « It is not straining the use of words to speak of this as the relation of probability »⁵.

Le rapport entre une croyance totale de type logique bivalente (le vrai ou le faux) et une croyance partielle de type probabiliste (située dans l'intervalle fermé $[0,1]$) est beaucoup plus problématique, car le premier type de croyance appartient à la logique formelle tandis que le second appartient à la « logique humaine », c'est-à-dire à l'induction. Or, pour Keynes, l'induction est une généralisation de la logique déductive⁶ ce qui n'est pas le cas pour Ramsey, en effet pour ce dernier l'induction doit être vue comme le *complément pragmatique* de la

¹ Wittgenstein (1922) voyait dans l'induction un processus qui ne pouvait relever que de la psychologie empirique et surtout pas de la logique, voir les propositions 6.31, 6.363 et 6.3631.

² Pour une bonne introduction, courte, à ces interprétations voir Gillies (2000), p 25-49, et p 88-112 ; Hacking (2004), p 153-164.

³ Pour Ramsey, le point de vue fréquentiste est pertinent lorsque l'on parle, par exemple, de la *proportion* de patients qui guérissent d'une maladie. Par ailleurs, l'auteur soutient que dans le cas où les probabilités seraient pensées comme étant une branche des mathématiques – et non de la logique comme c'est le cas dans l'article – alors l'interprétation la plus simple et la plus naturelle du point de vue du langage est l'interprétation fréquentiste. Voir Ramsey, op cit, p 159. Remarquons en passant que ce point de vue est celui de la plupart des manuels de mathématiques.

⁴ Ramsey, op. cit., p 156.

⁵ Keynes (1921), p 5-6.

⁶ Voir, Gillies (2000), p 30.

logique formelle : la dernière nous dit comment penser, la première sur quelle base agir¹.

Une autre différence majeure entre les deux auteurs se situe au niveau du concept de croyance : Keynes parle dans son texte de 1921 de « degré rationnel de croyance » (« rational degree of belief »)², c'est-à-dire du degré de croyance auquel tout homme doué de son bon sens doit parvenir avec une certaine quantité d'information. Mais cela va contre l'idée de Ramsey d'un degré de croyance *subjectif*, or pour Keynes, la relation de probabilité qui relie deux propositions entre elles ne peut-être subjective : « ...In the sense important to logic, probability is not subjective (...). A proposition is not probable because we think it so »³. Les différences se jouent donc à deux niveaux : dans la définition de la probabilité elle-même et dans la possibilité pour la mesure probabiliste d'être subjective. Bien qu'en désaccord avec Keynes, Ramsey retiendra pourtant de sa théorie le concept de « degré de croyance » et son rapport avec un certain type d'information comme nous allons le voir dans l'exemple qu'il développe. La critique que mène notre auteur porte surtout sur le rapport problématique qui est censé relier entre elles les propositions logiques. D'une part, Ramsey ne croit pas que le degré de croyance entre prémisses et conclusion soit transitif⁴, et, d'autre part il ne comprend pas comment ce rapport pourrait être perçu ou compris :

« Il ne me semble pas y avoir réellement de relations de probabilité telle qu'il les décrit. Il suppose que dans certains cas, celles-ci peuvent être perçues. (...) Je ne les perçois pas et s'il fallait qu'on me persuade qu'elles existent, il faudrait que ce soit par un argument »⁵.

Il n'y a chez Keynes aucun moyen objectif de mesurer (ou de quantifier) les degrés de probabilités qu'entretiennent entre elles les propositions. Keynes fait ici l'hypothèse d'une « capacité » qu'il ne définit pas vraiment de façon rigoureuse, une « capacité » presque obscure : « While we may possess a faculty of direct recognition of many relations of probability, as in the case of many other logical relations, yet some may be made more easily recognisable than others »⁶. Dans ce cadre-là, le rôle de la logique probabiliste serait de nous

¹ Comparer avec Savage (1972), p 59.

² Keynes, op. cit., p 4.

³ Ibid.

⁴ Voir, Ramsey, ibid.

⁵ Ramsey, op. cit., p 157.

⁶ Keynes, op. cit., p 53.

faire découvrir ces relations plus facilement. Ainsi que le fait Ramsey, nous pouvons mettre en doute cette “capacité” de percevoir les relations logiques, même si elle comporte des degrés et n’est pas équipotente chez tous les sujets¹. Par ailleurs, si cette faculté n’est pas partagée par tous, dans quelle mesure peut-elle être un guide pour mesurer les croyances et par là même nous disposer à l’action? Comme le remarque Ramsey : « C’est un peu comme si tout le monde connaissait les lois de la géométrie mais que personne ne pouvait dire d’un objet donné s’il est rond ou carré »².

D’autre part, et c’est le but de l’auteur dans cet article, il faut pouvoir quantifier les degrés de croyance, ce que Keynes ne fait pas³. Ramsey fait aussi remarquer, et c’est important, que lorsque nous tentons d’évaluer le degré de probabilité qui relie deux propositions nous tenons toujours compte de nos croyances personnelles : c’est ce que nous faisons dans la vie courante.

Il y a cependant un accord minimum entre les deux auteurs au moins sur un sujet que Ramsey souligne, c’est à propos du lien fort entre probabilités et rationalité⁴. Mais l’accord est de courte durée, car notre auteur n’accepte pas l’idée de Keynes selon laquelle un certain type de relations logiques – non définies par ailleurs – pourraient justifier un degré de croyance rationnel, et ce à la place de la relation de probabilité que Keynes proposait un peu plus haut comme explication. Il y aurait donc en fait deux types d’explications incompatibles entre elles comme le note Ramsey. Pour lui, le degré de croyance ne peut être justifié ni par une relation de type probabiliste indéfinissable entre deux ensembles de propositions, ni par une relation de type logique, non moins indéfinissable, qui viendrait – on ne sait ni pourquoi, ni comment – remplacer la relation probabiliste initiale. En effet, comme le dit Keynes lui-même, cité par Ramsey : « ...Nous ne pourrions jamais connaître le degré de probabilité qui serait justifié par la perception de relations logiques que nous sommes et serons à jamais incapables

¹ Voir, Gillies (2000), p 52.

² Ramsey, *ibid*.

³ Voir, Keynes, *op. cit.*, p 39. Les degrés intermédiaires entre la certitude et l’impossibilité sont des « probabilités non numériques ». On ne peut être plus loin du point de vue de Ramsey, et ce pour deux raisons : c’est bien les degrés de croyance intermédiaires entre 0 et 1 qu’il s’agira de formaliser, mais surtout de *mesurer*.

⁴ La citation que fait l’auteur est importante : « La probabilité est en un sens (...) relative aux principes de la raison humaine », *op. cit.*, p 158. C’est presque mots pour mots ce que Ramsey soutiendra un peu plus tard lorsqu’il dira que les probabilités doivent « être nécessairement vraies de n’importe quel ensemble *cohérent* de degré de croyance », *op. cit.*, p 174, c’est nous qui soulignons.

d’embrasser »¹. Il faut donc, et c’est ce à quoi va s’employer l’auteur immédiatement explorer une autre voie pour déterminer le degré de croyance.

Le but de Ramsey dans la partie suivante, « degrés de croyance »², qui s’articule autour de deux moments et qui est pour nous la partie la plus importante, va être d’établir « une méthode purement psychologique de mesure de la croyance »³. C’est à partir de cette mesure et de l’axiomatique que seront dérivées les probabilités. Pour parvenir à cette fin il faut tout d’abord postuler que nos croyances peuvent être mesurées. Mais ce postulat ne va pas de soi : ne dit-on pas que ce qui relève de la psychologie n’est pas mesurable⁴? Le problème qui émerge en premier lieu est celui-ci : comment mesurer un degré de croyance? Existe-t-il une échelle numérique des croyances? Nous savons en effet que les croyances sont susceptibles de degrés et d’intensité. Que dire dans ce cas de l’introspection? Il va falloir parvenir à subsumer tous ces éléments sous une unité de mesure commune⁵.

L’idée de Ramsey est de discuter les aspects d’une théorie de la mesure en partant du principe important que : « Le degré de croyance (...) n’a aucune signification à moins que nous spécifions de manière précise comment il doit être mesuré »⁶. C’est une idée centrale en ce sens qu’elle signifie que *c’est la mesure qui donne sens à ce qui est mesuré*. La mesure n’est donc pas une procédure arbitraire, même si son unité peut être établie de façon conventionnelle.

Nous retrouverons le problème de la mesure chez von Neumann/Morgenstern et Savage, problème lié au passage du qualitatif au quantitatif⁷. Les difficultés propres au concept de mesure sont aussi partagées par la physique contemporaine : précision et « ambiguïté dans la définition du processus de mesure »⁸. Le parallèle avec la physique contemporaine sera un des traits dominants de cette partie de l’article, la leçon à tirer de cette

¹ Ramsey, op. cit., p 159.

² Ramsey, op. cit., p 161.

³ Ibid. L’auteur se réfère, comme le feront après lui von Neumann et Morgenstern, au problème de la mesure en physique moderne. La théorie physique semble être dans les deux cas paradigmatique. Voir, von Neumann/Morgenstern (1947), p 20-24.

⁴ Les études pratiquées par Fechner (1860) avaient déjà solidement miné ce préjugé, dont on peut penser qu’il vient de Hegel. Par ailleurs les études menées par Edwards à partir des années 50 viendront finir de ruiner le préjugé en question, car une nouvelle discipline va alors émerger : la psychologie mathématique. Voir à ce sujet, Dawns/Tversky/Coombs (1970) ; voir aussi en français Dicks/Tournais/Flieler/Kop (1994).

⁵ Voir Picavet (1994), p 173 et suivantes.

⁶ Ramsey, op. cit., p 162.

⁷ Voir, Savage (1954), p 30-40 ; Granger (1967), p 171-174 ; et Granger (1988), p 255-261.

⁸ Ramsey, Ibid.

comparaison est que ce qui fonctionne en physique doit fonctionner en philosophie. Il faut donc établir un système à l'intérieur duquel n'importe quelle croyance puisse être mesurée de la façon la plus précise possible : une véritable échelle graduée des degrés de croyance. Il est clair qu'à la différence de Keynes cette échelle doit être numérique. Nous devons pouvoir quantifier une intensité¹, et cette quantification devra être une mesure probabiliste : « puisque nous calculons en fait des probabilités numériques »².

Pour parvenir à résoudre ce problème nous avons deux solutions :

(a) Nous placer du point de vue de l'intensité qui accompagne nos croyances, ce faisant « Par degré de croyance nous voulons dire l'intensité de ce sentiment (feeling) »³.

Mais, comme le soutiendra aussi Savage, la quantification de l'intensité d'une croyance par l'introspection ne va pas de soi, d'autant plus que comme chez Hume, il faut ici séparer les croyances des émotions. Par ailleurs nos croyances les plus fortes ne sont accompagnées d'aucun sentiment particulier⁴.

(b) Une deuxième possibilité, et c'est celle que retiendra Ramsey, consiste à voir dans la croyance une propriété causale⁵. C'est l'idée de la croyance comme cause de la disposition à l'action : dans quelle mesure serai-je prêt à agir sur la base de telle ou telle croyance⁶ ?

¹ C'est ce que Fechner avait pressenti dans son ouvrage de 1860. Nous y reviendrons, car la loi de Fechner-Weber aura une importance particulière lorsque nous parlerons du concept de « seuil psychologique », notamment avec l'article d'Allais (1953), mais aussi avec la théorie de la perspective (TP) de 1979.

² Ramsey, op. cit., 163.

³ Ibid.

⁴ Il faudrait peut-être nuancer et complexifier cette thèse. Nous nous y emploierons dans notre troisième chapitre. Le cas de la croyance en Dieu, par exemple, s'accompagne chez les sujets de divers sentiments. D'autre part il est peu probable que les croyances soient isolables les unes des autres et du tissu cognitif, notre troisième chapitre tendra vers un certain holisme de la croyance.

⁵ C'est la thèse pragmatique classique d'Alexander Bain (1859), qui sera reprise par Pierce (1907), Savage (1972) et discutée par Russell (1921). Cette thèse inspirera d'ailleurs Ramsey dans sa dernière partie.

⁶ Russell (1921) fait deux objections à cette thèse : d'abord, toutes les croyances ne mènent pas à l'action. Ramsey répond en proposant le concept de « croyance dispositionnelle ». Deuxièmement, comment peut-on distinguer les croyances des autres idées si il est impossible de les distinguer par leurs effets.

Cette thèse de la croyance comme disposition à l'action – mais seulement dans certaines circonstances précises¹ – va être discutée assez longuement par l'auteur. Ce qui nous intéresse ici c'est la question à laquelle il arrive ensuite : « Essayons maintenant de trouver une méthode de mesure des croyances en tant que bases d'actions possibles »². Nous pouvons dire que c'est ici à proprement parler que « commence » la théorie de Ramsey et que son analyse vient rejoindre de façon plus profonde notre interrogation sur l'axiomatisation de la décision et de l'utilité.

« La façon bien établie depuis longtemps de mesurer la croyance de quelqu'un est de proposer un pari, et de regarder quelle est la plus basse cote qu'il acceptera »³.

Pour déterminer le degré de croyance en une proposition « p » il faudra donc proposer un pari au sujet⁴. Car le pari a cet avantage qu'il est susceptible d'être quantifié facilement : si je mise à 3 contre 1 sur la proposition p : « il va pleuvoir demain », cela signifie que j'estime que $p = 0,75$ et non $p = 0,25$. Le pari n'est pourtant pas une méthode idéale comme le remarque Ramsey, car des sujets peuvent se comporter différemment face au même pari selon qu'ils seront ou non aversifs au risque⁵.

Par ailleurs il faudra, justement du fait de l'aversion au risque et de l'utilité marginale décroissante de l'argent, éviter les paris monétaires. Comme le note l'auteur, d'une façon générale : « Le fait de proposer un pari est susceptible de modifier l'état de son opinion »⁶. Le pari n'est donc pas un outil neutre pour la mesure des croyances, mais c'est une méthode « fondamentalement correcte »⁷.

« Je propose de prendre pour base une théorie psychologique générale (...) selon laquelle nous

¹ Voir, Ramsey, op. cit., p 164.

² Ramsey, op. cit., p 166.

³ Ibid.

⁴ Ici il convient de remarquer que le pari concerne une proposition, car c'est justement sur un type de proposition particulier que repose toute la construction de Ramsey. Savage utilisera lui aussi le concept de pari, par contre von Neumann/Morgenstern lui préféreront le concept de loterie d'où tout élément subjectif est banni.

⁵ Le concept d'aversion au risque est central en théorie de la décision, il avait été thématiqué par Bernoulli (1738). Ce concept est au fondement du paradoxe d'Allais et il joue un rôle capital dans le problème du renversement des préférences, voir Lichtenstein/Slovic (2006), mais aussi en théorie de la perspective. Une idée essentielle de notre troisième chapitre sera de connecter ce concept avec celui d'akrasie ou de faiblesse de la volonté.

⁶ Ibid.

⁷ Ibid.

agissons suivant la manière qui nous semble la plus propice à la réalisation de nos désirs, de sorte que les actions d'une personne sont entièrement déterminées par ses désirs et ses opinions »¹.

Nous reconnaissons ici la théorie de Hume selon laquelle nos désirs sont les causes de nos croyances et les moteurs de nos actions². Le caractère « naturel » de cette hypothèse pragmatique³ en fait un candidat théorique intéressant en ce sens qu'elle contient de façon virtuelle⁴ le concept de maximisation de l'utilité attendue (MUA). Par ailleurs il faut noter que cette théorie ne doit pas être confondue avec l'utilitarisme de Bentham dans laquelle « le plaisir occupe une place dominante »⁵. Il s'agit de *définir la croyance et le désir de façon opératoire*, c'est-à-dire comme conditions de l'explication de l'action, autrement dit : « Nous recherchons ce que nous désirons »⁶. Le « bien » auquel une personne aspire est défini en fonction du désir de cette personne pour le bien en question, et il ne faut voir ici aucune référence à un quelconque contenu éthique. Les concepts de « bon » et de « mauvais » sont utilisés de façon analogique « pour dénoter ce qui suscite chez une personne donnée le désir ou l'aversion »⁷. D'autre part, comme il s'agit d'augmenter une certaine quantité de bien en fonction d'un certain degré de croyance il faudra faire usage du principe de l'espérance mathématique. C'est là un des traits du bayésianisme de Ramsey, ce qui signifie que les processus décisionnels auront un aspect computationnel. Nous retrouverons ce trait chez von Neumann/Morgenstern et Savage⁸. Le degré de croyance est ainsi relié de façon directe à un principe mathématique, ce qui va permettre une première formalisation.

Dans la suite de son texte, notre auteur donne un exemple simple de la façon dont doit être déterminé le degré de croyance : il est équivalent à la différence entre la croyance totale de valeur arbitrairement fixée à 1 et le rapport du désavantage à acquérir de l'information

¹ Ramsey, op. cit., p 166-167.

² Hume, (1739/1746), p 462. La thèse de Hume est que la raison pratique serait l'instrument des passions et des désirs. Cette idée sera reprise dans la *Théorie des sentiments moraux* d'Adam Smith avec de légères modifications. Il s'agit là d'un des fondements philosophiques de l'utilitarisme.

³ Voir entre autres Claudine Tiercelin (1993) : *Peirce et le pragmatisme*. P.U.F.

⁴ Voir, Picavet (1994) p 64.

⁵ Ramsey, ibid. En ce qui concerne la théorie de Bentham on peut se reporter à l'article classique de Stigler (1950), p 309-311. La théorie de l'utilité que propose Bentham (1789) en faisant du bien ou de l'utile le synonyme du plaisir est discutée par de nombreux théoriciens de la décision.

⁶ Ibid.

⁷ Ibid.

⁸ Chez Ramsey si p est une proposition crue au degré m/n , il faudra alors multiplier les biens ou les maux pour lesquels la réalisation de p est nécessaire par ce rapport. Voir, Ramsey, op. cit., p 168.

(noté $f(x)$)¹ sur la différence entre l'avantage de ne pas se tromper (noté r , c'est-à-dire atteindre son but) et le désavantage de se tromper (noté w , c'est-à-dire ne pas atteindre son but) :

$$p = 1 - \frac{f(x)}{r - w}$$

Cette formule pourrait être explicitée comme exprimant la valeur actuarielle de la loterie où la probabilité p de gagner r et $(1-p)$ de perdre w est $(r-w)(1-p)$. Ramsey démontre cette formule en partant de la probabilité $p = m/n$ où n est le nombre de fois où on joue et m le nombre de fois où l'on gagne. Par là même est démontré que le degré de croyance oscillera entre 0 et 1, ce qui le met en rapport immédiatement avec le calcul des probabilités. Cette façon de mesurer le degré de croyance en une proposition a aussi la vertu de confirmer la méthode du pari : « En supposant que le pari soit en biens et en maux plutôt qu'en argent, il acceptera les paris dont la cote est préférable à celle correspondant à son état de croyance »². État de croyance qui aura été déterminé par la méthode ci-dessus. Notons d'autre part que c'est le degré de croyance en une proposition p qui va être déterminant, c'est-à-dire le degré de croyance selon lequel p est vrai ou non p est vrai. Une question émerge ici : comment le degré de croyance en p peut-il conditionner la disposition du sujet en faveur d'une certaine décision ? Il convient aussi de se demander dans quelles limites le degré de croyance en p peut-il avoir un impact sur les préférences des sujets et leur ordonnancement ?

3) Les propositions éthiquement neutres et l'axiomatique.

Nous devons tout d'abord partir du principe que le sujet a des croyances à propos de toutes les options possibles³ et que ses croyances arrivent par degrés. Cela nous permettrait, en faisant l'hypothèse que nous soyons omnipotents et omniscients, et en lui proposant des paris de découvrir comment il ordonne toutes les options possibles et imaginables : « De cette

¹ La formalisation de l'acquisition de l'information en vue de la modification des croyances est un trait fondamental du bayesianisme. Voir Savage (1954), p 43-50 ; Hacking (2004), p 185-198 ; ainsi que Oaksford/Chater (2006).

² Ramsey, op. cit., p 169.

³ Ce principe, peu réaliste est partagé par von Neumann/Morgenstern (1947) et par Savage (1972).

manière, tous les mondes possibles seraient placés par ordre de valeur »¹. Deux remarques s'imposent : tout d'abord l'emploi que fait Ramsey du concept de « monde possible ». Et ensuite l'auteur ne propose pas la distinction canonique dans le domaine de la décision en contexte d'incertitude², ou plutôt qui deviendra telle à partir de Savage, entre actes, états du monde et conséquences³.

L'idée de « monde possible », très connotée philosophiquement depuis Leibniz (1714) et logiquement depuis sa reprise par Kripke (1962), (1963), et Lewis (1986), correspond chez Ramsey aux entités ultimes entre lesquelles le sujet doit choisir, il dit que ce sont les « différentes totalités possibles d'événements »⁴. Ce sont les termes des alternatives elles-mêmes, les perspectives offertes par les paris, les objets du choix rationnel. Du point de vue formel, Ramsey ne fait pas différence entre un monde possible, une option, et la valeur de monde possible ou de cette option, il les note ainsi : α , β , γ , δ ... Un monde possible est équivalent à sa valeur⁵, il n'intervient dans la théorie que par la valeur de son option.

Ce dont nous avons besoin maintenant c'est d'être capable de déterminer les « différences de valeurs » (values differences) entre les mondes possibles pour pouvoir les comparer. C'est ici que vont intervenir les PEN crues au degré $\frac{1}{2}$. Une situation décisionnelle type peut-être énoncée comme suit : « Préfereriez-vous un monde α dans tous les cas, ou le monde β si p est vraie et le monde γ si p est fausse »⁶. La valeur de vérité de p entre ici comme *condition*, et nous pouvons déjà voir que selon le degré de croyance attribué à p le choix risque d'être modifié. Pour trouver un point d'indifférence il faudra donc que p soit crue au degré $\frac{1}{2}$ ⁷. Par ailleurs, afin de pouvoir jouer leur rôle de marqueur du point d'indifférence, les propositions devront être aussi éthiquement neutres.

Qu'est-ce qu'une PEN ? « Une proposition atomique p est appelée éthiquement neutre si deux mondes possibles ne différant que quant à la vérité de p sont toujours d'égale valeur »⁸. Une PEN sera donc telle qu'elle n'affectera pas les préférences et leur

¹ Ramsey, op. cit., p 170.

² La distinction entre les différents contextes a été proposée par Knight en 1921. Nous y reviendrons dans nos deuxième et troisième parties.

³ Voir, Savage (1954), p 6-7 ; Luce/Raiffa (1957), p 277 ; Chernoff (1959), p 119-167 ; Fishburn (1970), p 163-170, et p 191-210 ; Kreps (1988), p 127 et suivantes ; enfin voir aussi Binmore (2009), p 2 et suivantes.

⁴ Ramsey, *ibid*.

⁵ Voir, Picavet (1994), p183-184.

⁶ *Ibid*.

⁷ Voir, Sahlin (1990), p 20-23 ; Jeffrey (1965/1983), p 46-49 ; Picavet (1996), p 185-186.

⁸ Ramsey, *Ibid*.

ordonnancement par le sujet, il faut donc qu'elle ne connote en fait rien de spécial de type informationnel qui pourrait influencer le choix¹. Bien entendu ce type de proposition ne doit avoir aucune référence morale ou culturelle particulière, car cela fausserait immédiatement la donne. Ce sont donc des propositions assez rares, et il est certain qu'elles ne font pas partie de celles que nous employons le plus souvent dans nos conversations. Par exemple des propositions telles que : « le nombre de cheveux sur ma tête est supérieur à 100000 » ou « le nombre d'atomes dans l'univers est pair » peuvent être caractérisées comme des PEN, dans une situation décisionnelle donnée elles ne peuvent altérer les préférences du sujet.

Comme nous le remarquons plus haut, les PEN sont toutes assez rares et peu utilisées, et il convient de se demander si elles sont réellement opératoires ? Afin qu'elles puissent le devenir pour déterminer un point d'indifférence, la condition *sine qua non*, est qu'elles soient l'objet d'une croyance de $\frac{1}{2}$. Or :

« Le sujet est dit avoir une croyance de degré $\frac{1}{2}$ en une telle proposition p s'il n'a aucune préférence entre les options (1) α si p est vraie, β si p est fausse, et (2) α si p est fausse, β si p est vraie mais a une préférence simple entre α et β »².

Ce qui d'un point de vue plus formel peut-être traduit comme suit : pour tout α et pour tout β inclus dans l'ensemble des états du monde, pour tout p compris entre 0 et 1, il est vrai que :

$$[(\alpha \succ \beta) \vee (\beta \succ \alpha)] \supset \{[(\alpha \mid p) \vee (\beta \mid \neg p)] \sim [(\alpha \mid \neg p) \vee (\beta \mid p)]\}$$

Que nous indique cette formule ? Il sera peut-être plus éclairant de prendre un exemple. Prenons deux mondes possibles α et β , α = il fera beau demain, β = il pleuvra demain, et une PEN crue au degré $\frac{1}{2}$ soit $p = 1$ ou 2 ou 3 est produit au prochain lancé d'un dé à six faces non truqué. Nous savons par ailleurs que le sujet préfère α à β . La croyance de degré $\frac{1}{2}$ en p est ce qui garantit l'indifférence entre les deux options : en effet, mes chances de « gagner » ou de « perdre » sont équivalentes dans les deux options. L'espérance

¹ Il y aurait des comparaisons à mener entre les PEN et le PCS de Savage dans lequel c'est la réalisation ou pas d'un événement de référence qui ne doit pas avoir d'influence sur les préférences.

² Ramsey, Ibid.

mathématique de α – puisqu’on ne peut pas encore parler d’utilité – est la même dans les deux membres qui suivent l’implication. Imaginons maintenant, que toujours avec les mêmes préférences, nous ayons une croyance différente en p , disons 1/3 et par définition 2/3 pour non p . Dans ce cas-là il est bien évident que c’est la seconde alternative qui devra être préférée puisqu’elle favorise le monde possible que le sujet préfère. Une préférence inverse amènerait à préférer le membre de droite.

Les PEN jouent donc le rôle du marqueur d’indifférence entre les deux alternatives : « Indifférence entre parier d’un côté ou de l’autre pour les mêmes enjeux »¹. C’est le premier moment dans le passage du qualitatif vers le quantitatif : *trouver une unité de mesure*. Les PEN vont donc nous servir non seulement à calibrer un point d’indifférence entre deux mondes possibles représentés par leurs valeurs, mais aussi à évaluer les différences de valeurs entre les alternatives. C’est-à-dire que nous pouvons savoir comment le sujet classe ses préférences de façon ordinale, mais aussi comment les différentes options sont évaluées de façon cardinale. Cependant il manque encore à Ramsey un système d’axiomes pour que sa théorie soit encore plus précise.

Notre auteur va ensuite expliquer comment, grâce aux PEN, nous pouvons déterminer l’équivalence des différences de valeurs entre deux mondes possibles dans deux paris différents. Soit, par exemple, α et β d’une part, γ et δ d’autre part : « Si p est une PEN qui est crue au degré 1/2, le sujet n’a pas de préférence entre les options (1) α si p est vraie, δ si p est fausse, et (2) β si p est vraie, γ si p est fausse »². Ce qui formellement équivaut à : pour tout α et pour tout β pour tout γ inclus dans l’ensemble des états du monde, pour tout p compris entre 0 et 1, nous avons la définition suivante :

$$(\alpha\beta = \gamma\delta) \equiv \{[(\alpha \mid p) \wedge (\delta \mid \neg p)] \sim [(\beta \mid p) \wedge (\gamma \mid \neg p)]\}$$

Les deux équations que nous venons de mentionner sont ce qui constitue le fondement du système de mesure de Ramsey, système qui s’applique aux options et à leurs différences. Ce système devra aussi bien entendu respecter la transitivité des relations « = » et « > » entre

¹ Ramsey, op. cit., p 171.

² Ibid.

les préférences afin de pouvoir les ordonner¹. Mais ces deux équations resteraient lettre morte si notre auteur n'y adjoignait une axiomatique du choix rationnel. Les 8 axiomes que nous allons étudier, interprétés comme condition de possibilité des décisions rationnelles, ont un aspect normatif. On peut distinguer trois groupes : (1)-(4), (5)-(6) et (7)-(8).

« (1) Il y a une proposition éthiquement neutre crue au degré $\frac{1}{2}$. »².

Cet axiome qui a une certaine implication ontologique garantit l'existence d'une PEN, car bien que Ramsey ne fasse pas mention du problème ici, le statut des PEN pose de sérieux problèmes. C'est toutefois l'introduction du concept de PEN qui donne à la théorie de Ramsey son originalité indiscutable. Jeffrey (1965/1983) réactualisera les vues de Ramsey³ afin de montrer comment nous pouvons passer de la désirabilité de l'agent pour une option aux probabilités qu'il attribue à cette option⁴. De leur côté, von Neumann/Morgenstern et Savage ne reprendront pas la théorie des PEN, et c'est de façon nouvelle qu'ils établiront le point d'indifférence dans leurs axiomatiques, tout en restant dans le cadre d'une théorie de la mesure. Dans la théorie de Ramsey, les PEN sont des prérequis opérationnels⁵, puisque c'est à partir de ce type de propositions que l'on peut établir une première différence entre les options et les paires d'options. Les PEN fixent les unités de mesure de l'échelle subjective.

« (2) si p , q sont de telles propositions et que l'option α si p , δ si non- p est équivalente à β si p , γ si non- p , alors α si q , δ si non- q est équivalent à β si q , γ si non- q »⁶.

L'axiome (2) concerne l'applicabilité des PEN aux options et l'interchangeabilité des PEN entre elles pour des options équivalentes. On peut donc utiliser différentes PEN pour calibrer les points d'indifférence entre les options qui elles restent les mêmes. Autrement dit, le fait de passer d'une PEN $\frac{1}{2} p$ à une PEN $\frac{1}{2} q$ n'altère pas l'équivalence entre les différentes

¹ Comme nous allons le voir sans tarder, à la différence de von Neumann/Morgenstern (1947) et Savage (1954), il n'y a pas chez Ramsey d'axiome d'ordonnancement qui soit clairement posé. Peut-être l'auteur avait-il à l'idée que cette notion allait de soi ?

² Ramsey, Ibid.

³ Notamment en introduisant des PEN de valeurs différentes, ce que Ramsey n'admet pas dans son axiomatique. Dans l'optique de Jeffrey, l'introduction de PEN de différentes valeurs permet d'affiner l'échelle des préférences subjectives.

⁴ Voir, Jeffrey, op. cit., p 49-50.

⁵ Voir, Sahlin (1990), p 28.

⁶ Ramsey, ibid.

options, et elles permettent d'établir l'égalité suivante : $\alpha\beta = \gamma\delta$ ¹. Cet axiome signifie donc que si les options sont équivalentes eut égard à des PEN différentes alors nous pouvons dire que la différence de valeur entre les options constitutives des alternatives sont aussi équivalentes².

(2a) est une conséquence directe de (2) :

« Si $\alpha\beta = \gamma\delta$, alors $\alpha > \beta$ est équivalent à $\gamma > \delta$ et $\alpha = \beta$ est équivalent à $\gamma = \delta$ »³.

Si les différences de valeurs entre les options sont égales, alors les relations de préférence ou d'indifférence entre les options elles-mêmes doivent aussi être équivalentes. Il s'agit là aussi d'un axiome d'ordonnement⁴.

« (3) Si l'option A est équivalente à l'option B et B à C , alors A [est équivalente] à C »⁵.

C'est un axiome classique de transitivité entre les options ou entre les paris sur les options. Remarquons ici que le vocabulaire de Ramsey est plutôt flottant et relâché⁶, car il n'a jamais fait usage des lettres majuscules romaines avant, et il ne définit nulle part à quelles entités décisionnelles elles peuvent correspondre. Nous postulons donc qu'il s'agit d'une métavariable représentant à la fois les options et les paris. L'axiome de transitivité est fondamental, car il garantit la cohérence des préférences⁷.

¹ Cela veut-il dire que toutes les PEN sont équivalentes ? Probablement. Car sinon, il faudrait envisager la possibilité qu'il y ait au moins deux types de PEN différentes ce qui n'est mentionné nulle part par Ramsey. Par contre, quand nous complexifions l'échelle des préférences, les PEN peuvent ne pas toutes être crues au degré $\frac{1}{2}$, elles peuvent recevoir toutes les valeurs comprises entre 0 et 1.

² Ce qui permet à Ramsey de « déduire » les théorèmes suivants : « si $\alpha\beta = \gamma\delta$, alors $\beta\alpha = \delta\gamma$, $\alpha\gamma = \beta\delta$, $\gamma\alpha = \delta\beta$ », *ibid.* Ce qui veut dire que les valeurs sont commutatives et substituables. À partir du moment où l'égalité entre les différences de valeurs est acquise, l'agent peut ranger les options de n'importe quelle façon.

³ Ramsey, *Ibid.*

⁴ Ici « > » est à lire comme « préféré à ». Voir, Picavet (1994), p 187.

⁵ Ramsey, *ibid.*

⁶ Il faut garder à l'esprit qu'il y a bon nombre d'éléments seulement ébauchés ou indiqués dans cet article et qui ne recevront ni développement particulier, ni traitement systématique. Ce sera par exemple le cas de la fonction d'utilité et du théorème de la représentation.

⁷ Voir, entre autres, von Neumann/Morgenstern (1947), p 26 ; Herstein/Milnor (1953), p 292-293 ; Savage (1954), p 19-21 ; Luce/Raiffa (1957), p 25 ; Barberà/Hammond/Seidl (2004), p 690 et suivantes ; Binmore (2009), p 10 ; Peterson (2009), p 165-169. C'est un axiome essentiel, il sera cependant remis en question par Edwards (1954), et Slovic/Lichtenstein (1968). Il semble que les théories « classiques » de la décision utilisent toutes cet axiome. Cependant, certains comme Sahlin (1990) ainsi que Sugden/Loomes (1982) soutiennent que la transitivité n'est pas toujours nécessaire, et que dans certaines conditions comme la théorie du regret, un ensemble de préférences peut être non transitif sans que cela soit pour autant irrationnel. Voir, Picavet (1996), p 222 et suivantes.

En effet si je ne respecte pas ce principe et que j'ai un ensemble de préférences du type : $A > B$, $B > C$, $C > A$, alors, en imaginant que je suis prêt à miser une somme aussi petite soit-elle sur A ou B ou C , je perdrais de l'argent de façon systématique. Car imaginons que je mise 1\$ sur A , je serai prêt à miser un peu plus pour avoir C puisque C est préféré, mais je serai aussi prêt à miser un peu plus sur B , et donc encore un peu plus sur A , etc. C'est l'argument de la « pompe à fric » (« money pump ») il vient renforcer l'idée que la transitivité est la condition si ne qua non de la cohérence¹. L'axiome de transitivité a pour corollaire immédiat : « Si $\alpha\beta = \gamma\delta$ et $\beta\eta = \zeta\gamma$, alors $\alpha\eta = \zeta\delta$ »². Ce qui signifie que la transitivité fonctionne aussi à l'intérieur des options, c'est-à-dire lorsque nous « descendons » des paris aux valeurs des mondes possibles qui les composent.

« (4) Si $\alpha\beta = \gamma\delta$, $\gamma\delta = \eta\zeta$, alors $\alpha\beta = \eta\zeta$ »³.

Cet axiome signifie la même chose que le précédent, mais cette fois-ci concernant la *différence de valeur* entre les options. Ramsey veut s'assurer que la transitivité fonctionne à tous les niveaux, ce qui renforce le degré de cohérence des décisions.

Après ce premier groupe d'axiomes nous sortons de l'ordonnancement et de la transitivité. Puisque ce que nous garantissent les axiomes (1)-(4), outre l'existence d'une PEN $\frac{1}{2}$, c'est que nous pouvons ordonner nos préférences de façon rationnelle et cohérente.

« (5) $(\alpha, \beta, \gamma) \in (x) (\alpha x = \beta\gamma)$ »⁴.

Pour tout monde possible α, β ou γ il existe un résultat (x) unique telle que la différence de valeur entre α et x soit équivalente à celle entre β et γ .

Cet axiome permet de rendre *commensurable* une alternative quelconque avec une différence de valeur quelconque entre deux options du fait de l'existence d'un résultat (x) unique. Cet axiome assure ainsi que tout monde possible peut devenir le membre d'une

¹ Voir, Peterson (2009), p 167-168.

² Ramsey (1926).

³ Ibid.

⁴ Ibid.

alternative, et par là même être comparé.

« (6) $(\alpha, \beta) \in I(x) (\alpha x = \beta x)$ »¹.

Dans la même veine que l'axiome (5) celui-ci garantit tout simplement qu'entre deux alternatives il existe toujours un troisième résultat unique dont la valeur est exactement située à mi-distance entre les valeurs des deux premiers résultats. C'est un axiome de continuité.

Les axiomes (5) et (6) permettent d'établir l'existence de préférences intermédiaires entre les extrêmes 0 et 1 d'une part, et entre toutes les alternatives possibles sur l'échelle du sujet d'autre part. Ces axiomes nous permettent d'affiner, de façon indéfinie, les mesures des degrés de croyances et des préférences. Nos préférences sont certes organisées de façon ordinale (du fait des axiomes (1)-(4)), mais il y a aussi une détermination cardinale, due à l'action combinée de (5) et (6). Cette détermination cardinale permet de savoir *de combien* les alternatives rangées ordinalement sont préférées².

« (7) Axiome de continuité : toute progression a une limite (ordinal) »³.

Cet axiome est énoncé de façon non formelle, mais il est très important, car il garantit que l'ensemble des préférences doit avoir une limite inférieure et supérieure incluse dans l'ensemble. Sans cette clause nous pourrions nous retrouver face aux mêmes problèmes que ceux induits par les paradoxes de Zénon : nous ne pourrions jamais atteindre les degrés 0 ou 1 mais seulement nous en rapprocher indéfiniment⁴.

« (8) Axiome d'Archimède »⁵.

Cet axiome est juste mentionné et nous ne savons de façon exacte ce que Ramsey avait précisément à l'esprit. Il semblerait pourtant qu'il y ait au moins deux façons de considérer

¹ Ramsey (1926).

² Voir, Jeffrey (1965/1983), p 46-58 ; Sahlin (1990), p 30-34 donne un exemple intéressant de la façon dont tous les axiomes fonctionnent de façon combinée.

³ Ramsey, *ibid.*

⁴ Sur les liens entre les paradoxes de Zénon et la théorie des ensembles on peut consulter Tiles (1989), p 12-22.

⁵ Ramsey, *ibid.*

cette clause qui se répondent l'une à l'autre. Nous présentons la première interprétation à titre d'hypothèse, la seconde est due à Picavet (1994).

a) C'est un axiome de continuité qui pourrait signifier que pour une différence de valeur donnée, disons $\alpha\beta$, aussi petite que l'on veut, et une seconde différence de valeur, disons $\gamma\delta$, aussi grande que l'on veut, il existe un n entier tel que le produit de la plus petite différence de valeur par n deviendrait *supérieur ou égal* à la plus grande différence de valeur. Ou, de façon formelle :

$$(\alpha, \beta, \gamma, \delta)(\exists n)[(\alpha\beta > \delta\gamma) \supset (n(\alpha\beta) \geq \delta\gamma)]$$

Ce qui voudrait dire que toutes les différences de valeurs sont *commensurables* du fait que nous pouvons établir que :

$$n(\alpha\beta) = \delta\gamma$$

n étant ici l'unité de mesure entre les différences de valeurs¹.

b) La deuxième interprétation est proposée par Picavet (1994) qui s'inspire de Nozick (1985). Il s'agirait de faire de l'axiome (8) une clause de continuité, mais à la différence de Nozick et de la théorie de l'utilité, qui postulent une continuité entre les préférences et les distributions de probabilités², Picavet remplace les probabilités par une PEN. Cet axiome préfigurerait de façon lointaine celui que l'on trouve en théorie « classique » de l'utilité³. Picavet propose de formaliser l'axiome comme suit : « Si $\alpha > \beta$ et $\beta > \gamma$, il existe p telle que le sujet soit indifférent entre les options β et [α si p ; γ si non p]⁴. Ce qui signifie qu'il existe une PEN telle qu'elle rendrait le sujet indifférent entre l'option β certaine et deux autres options conçues comme les alternatives d'un pari. Remarquons que dans ce cas là nous passons d'une préférence stricte à une indifférence, l'introduction de p vient en quelque sorte neutraliser la

¹ La propriété archimédienne permet d'établir la supériorité ou l'égalité de tout x aussi petit soit-il avec tout y aussi grand soit-il, via un n entier. C'est un axiome communément utilisé en théorie de la mesure. Dans le cadre de son axiomatique, Ramsey avait probablement à l'idée qu'on puisse tirer de cette propriété une égalité.

² Voir, Herstein/Milnor (1953), p 293.

³ Voir, von Neumann/Morgenstern (1947), p 26 ; Herstein/Milnor (1953), p 292-293 ; Barbara/Hammond/Seidl (2004), p 690 ; Peterson (2009), p 100.

⁴ Picavet, op. cit., p187.

préférence originelle.

Selon l'auteur à partir de ces 8 axiomes une mesure d'utilité est dérivable : « Ces axiomes permettent aux valeurs d'être corrélées de façon biunivoque avec les nombres réels de sorte que si α^1 correspond à α etc., $(\alpha\beta = \gamma\delta) \equiv [(\alpha^1 - \beta^1) = (\gamma^1 - \delta^1)] \gg^1$.

Remarquons que le passage des différences de valeurs entre les mondes possibles vers l'utilité est tout juste mentionné par l'auteur qui d'ailleurs n'emploie nulle part dans son article le concept d'utilité ! Le seul indice de la possibilité d'établir une fonction d'utilité est l'équation susmentionnée qui signifie que l'égalité entre deux différences de valeurs est équivalente à l'égalité de la différence entre l'utilité de α (ici α^1) et l'utilité de β (ici β^1) d'une part et la différence entre l'utilité de γ (ici γ^1) et l'utilité de δ (ici δ^1) d'autre part. Ce qui nous permet d'établir une correspondance directe entre toutes les différences de valeurs et les utilités (U), du type :

$$(\alpha\beta = \gamma\delta) \equiv [(U(\alpha - \beta)) = (U(\gamma - \delta))]$$

Il faut noter que nous allons toujours des préférences vers l'utilité et non l'inverse². Ce qui intéresse l'auteur ici c'est de pouvoir dériver, à partir de la mesure des degrés de croyances le degré de probabilité subjective qu'un sujet entretient concernant une option ou un monde possible. Et c'est à cette dérivation originale que Ramsey va maintenant s'appliquer³.

Cette dérivation est en soi assez simple : la probabilité que le sujet attribue à ($P(p)$) sera équivalente à la cote à laquelle « il parierai maintenant sur p »⁴. D'une façon plus formelle : si

$$\alpha \sim [(\beta \mid p) \vee (\gamma \mid \neg p)]$$

La représentation probabiliste du degré de croyance en p sera :

¹ Ramsey, op. cit., 172.

² Voir Luce/Raiffa (1957), p 31.

³ Nous retrouvons ce processus chez Savage et Jeffrey, ainsi que chez de Finetti : c'est l'essence même du subjectivisme que de parvenir à extraire une mesure probabiliste à partir d'une mesure de croyance. Voir, Jeffrey (1965/1983), p 55-58, Kyburg/Smokler (1980) pour les textes fondateurs ; Gillies (2000), p 50 et suivantes.

⁴ Ramsey (1926).

$$P(p) = \frac{\alpha - \gamma}{\beta - \gamma}$$

C'est-à-dire qu'il sera égal au rapport de la différence des options entre elles. Degré de croyance et probabilité sont ainsi directement liés : la probabilité nous indique la mesure quantitative du degré de croyance. Comme les axiomes nous permettent d'ordonner, d'affiner et de limiter nos préférences entre les degrés de croyances nous pouvons obtenir une échelle de probabilités subjectives aussi fine que nous voulons¹. À partir de là, Ramsey est aussi en mesure de définir le concept central de probabilité conditionnelle² : « Le degré de croyance en p étant donné q »³, que l'auteur prend soin de distinguer de l'implication logique ($p \supset q$) et de certains problèmes liés à la fermeture sous l'implication⁴. Il s'agit ici de savoir la cote à laquelle le sujet parierait sur p à la condition que q ⁵. La méthode de dérivation est la même que pour les probabilités non conditionnelles : si

$$[(\alpha | q) \vee (\beta | \neg q)] \sim [(\gamma | p \wedge q) \vee (\delta | \neg p \wedge q) \vee (\beta | \neg q)]$$

Alors la représentation probabiliste du degré de croyance en $p|q$ sera, en éliminant $(\beta | \neg q)$ des deux côtés de la relation :

$$p | q = \frac{\alpha - \delta}{\gamma - \delta}$$

Ici aussi la probabilité est équivalente au rapport de la différence entre les options « qu'on doit supposer être [le] même pour n'importe quel $\alpha, \beta, \gamma, \delta$ qui satisfont les conditions données »⁶. Il faut noter que le degré de croyance en q doit être inférieur à 1 car sinon la

¹ Voir, Jeffrey (1965/1983), p 50-53.

² Voir Hacking (2004), p 57-66.

³ Ramsey (1926).

⁴ Pour une introduction à cette problématique voir la quatrième partie du volume édités par Sosa/Kim/Fantl/McGrath (2009), p 231-302.

⁵ Sur le concept de conditionalisation, on peut se reporter à Stalnaker (1970), Lewis (1976), Oaksford/Chater (2006) et (2008) ; Lepage/Morgan (2011).

⁶ Ramsey, *ibid.*

certitude de q pourrait altérer complètement l'ensemble du système de croyances¹.

Il faudra ensuite s'assurer, ce que Ramsey ne fait que très partiellement « Qu'il est possible de prouver les lois fondamentales de la croyance probable (des degrés de croyances qui se situent entre 0 et 1) »². Il s'agit en fait des axiomes des probabilités, que l'on peut synthétiser comme suit³ :

- 1) $P(p) \geq 0$.
- 2) $P(T) = 1$.
- 3) Si p et q sont incompatibles : $P(p \vee q) = P(p) + P(q)$.

« Ce sont là les lois des probabilités, dont nous avons montré qu'elles sont nécessairement vraies de n'importe quel ensemble cohérent de degrés de croyances »⁴. *Les lois des probabilités sont ainsi érigées en lois de la cohérence et de la rationalité*. C'est une thèse très importante, car elle vient compléter la thèse traditionnelle de la logique comme constitutive des lois de la pensée : les probabilités seront les lois du choix rationnel. Cette séparation entre deux formes de logiques va alimenter les parties 4 et 5 de l'article. Les lois de la cohérence décisionnelle sont de deux types : ordonnancement et transitivité d'une part, et lois des probabilités d'autre part.

Comme le souligne Ramsey, la violation de la transitivité des préférences ferait dépendre le choix des sujets « de la forme précise sous laquelle les options lui seraient offertes, ce qui est absurde »⁵. En fait ce que Ramsey établit ici de façon non formelle c'est le principe d'indépendance qui fera couler beaucoup d'encre dès sa formulation – au demeurant implicite⁶ – par von Neumann/Morgenstern (1947) : les préférences doivent être indépendantes des

¹ Nous avons ici une anticipation très nette de ce que peut causer l'« effet certitude » dont la prospect theory montrera l'influence qu'il a dans le choix des sujets, voir Kahneman/Tversky (1979), p 20-22. Cet effet sera aussi largement exploité par Allais (1953) dans son « paradoxe ». Nous analyserons cet effet de près dans notre deuxième partie, car c'est justement lui qui va jouer contre l'axiome d'indépendance.

² Ramsey, op. cit., p 173.

³ Voir, Kolmogorov (1933).

⁴ Ramsey, op. cit., p 174.

⁵ Ibid. C'est précisément ce que la théorie de la perspective mettra en avant avec l'« effet de cadrage », voir Kahneman/Tversky (1986). Cet effet était déjà connu d'Allais (1953), et ses implications vont être développées de façon dramatique pour la TUA par Edwards (1954), mais surtout par Lichtenstein et Slovic (1968), (1971) et (1973), ainsi que Tversky/Kahneman (1971). Cet effet nous plonge au cœur du problème de la représentation et son influence déterminante dans le domaine de nos inductions habituelles et de nos décisions. Le problème général de la représentation est crucial pour penser le concept de décision.

⁶ C'est Malinvaud qui explicitera la clause à partir de l'article de Samuelson (1952).

probabilités¹ et de la forme dans laquelle les options sont présentées². Ramsey a ici à l'esprit le deuxième terme de la conjonction. Cet axiome s'énonce formellement comme suit :

$$(\alpha \succ \beta) \supset [(\alpha, p) + (\gamma, 1-p) \succ (\beta, p) + (\gamma, 1-p)]^3$$

Par ailleurs, si la loi d'additivité à 1 n'est pas respectée, le sujet pourrait être victime d'un « dutch book », c'est-à-dire d'un pari – proposé par un bookmaker habile – dans lequel, quelque soit la perspective il perdrait de façon systématique. Un *Dutch book*⁴ advient dans des situations du type suivant : soit la proposition p : « Malakov gagnera la prochaine course ». Un joueur inattentif aux axiomes des probabilités peut attribuer au cheval les chances suivantes : $P(p) = 0.6$, et $P(\neg p) = 0.2$. Imaginons un bookmaker habile qui soit par contre lui au courant des axiomes des probabilités, il pourra proposer à notre joueur les paris suivants :

- a) Gagner 6 \$ si $\neg p$ et perdre 4\$ si p .
- b) Perdre 8 \$ si $\neg p$ et recevoir 2 \$ si p .

Il est bien entendu que le bookmaker connaît les probabilités que le joueur attribut aux résultats, le pari a) aura donc une utilité subjective de : $(-4\$ \times 0,6) + (6\$ \times 0,4) = 0$. Le pari est donc équitable, le pari b) aura lui aussi une utilité attendue de 0. Si les deux paris sont présentés au joueur comme une paire il fera une perte sèche. En effet si Malakov gagne, le pari a) produira une perte de 4\$ et b) un gain de 2\$, donc une perte totale de 2\$; par ailleurs si Malakov perd, a) produira un gain de 6\$ et b) une perte de 8\$, ce qui fait encore une perte totale de 2 \$. Pour éviter ce cas de figure il faut que les assignations de probabilités soient cohérentes : si le joueur pense que Malakov a 0,6 chances de gagner, il doit avoir de façon

¹ Voir, Gärdenfors/Sahlin (1988), p 5 ; voir aussi Fishburn/Wakker (1995).

² C'est ce que les expériences de Lichtenstein et Slovic en 1971 et 1973 viendront réfuter, ainsi que celles de Kahneman et Tversky dans les années 70. Ellsberg (1961) avait déjà nettement anticipé ce problème.

³ Nous reviendrons abondamment sur cet axiome qui fit problème dès sa première formulation, comme l'attestent les remarques de Marschak (1950), de Samuelson (1952), Malinvaud (1952), Herstein/Milnor (1953), Allais (1953), Edwards (1954). C'est le point faible de la TUA concernant les réfutations empiriques. Avec Savage (1954) ce principe deviendra le PCS. C'est précisément sur ce principe et sa formulation dans Samuelson (1952) que porteront les critiques d'Allais et plus tard celles de Kahneman et Tversky.

⁴ Sur ce concept et ses différents types la littérature abonde, mais on peut se reporter à Peterson (2009), p 153-156 ; Binmore (2009), p 123. Le ressort du Dutch book réside dans le fait que le sujet considère deux éléments (p ou non- p) comme sous-additifs, c'est-à-dire dont les probabilités n'arrivent pas à 1 si on les additionne.

nécessaire 0,4 chances de perdre¹.

La logique probabiliste est en fait une logique inductive et une logique de la cohérence. Si un sujet a un ensemble de préférences cohérent concernant une proposition p , alors il doit être prêt à parier sur p « pour n'importe quelle mise, les mises étant mesurées en termes de valeurs ultimes »². Mais justement la question pourrait être : jusqu'où peut s'étendre l'application de la méthode du pari en contexte décisionnel ? Et Ramsey de répondre : « Il semble que toute notre vie nous ne faisons, en un sens, que parier »³. Car toute décision peut avoir des conséquences incertaines. Mais cette thèse n'est-elle pas un peu trop forte ? Car elle pourrait aussi bien signifier que dans nos existences, si nous parions tout le temps c'est que nous n'avons au fond aucune certitude. La question reste à débattre, puisque dans le domaine du certain il n'y a pas de place pour les paris, comme Ramsey le remarque lui-même.

L'originalité de Ramsey ici est d'intégrer le pari, comme méthode opératoire, à la rationalité de l'agent, via le concept de cohérence. Tout se tient : les axiomes garantissent la cohérence décisionnelle qui doit induire le sujet à parier. Mais il faut noter que le pari est aussi au fondement de l'axiomatique, intégré, pourrions-nous dire, au concept de PEN. Par ailleurs, comme la cohérence est synonyme de rationalité, le pari l'est aussi.

Dans la fin de cette partie, Ramsey revient sur l'idée que les paris ne doivent pas être pensés en termes monétaires. Mais il insiste surtout, et ce sera très important pour le reste de notre travail, à propos de la limitation des options en contexte décisionnel : « ...Rien n'a été dit lorsque le nombre des options possibles est infini »⁴.

Le problème soulevé par Ramsey anticipe ce que nous allons trouver chez Savage avec le concept de « petit monde » (« small world »)⁵. Ces remarques nous conduisent aussi à penser qu'il y a un aspect *limitatif* dans le domaine des théories de la décision. Ce que nous entendons ici par limitation ce sont deux choses : tout d'abord il est bien évident que nos alternatives réelles sont toujours limitées du point de vue spatio-temporel, et il en va de même

¹ Le Dutch book que nous présentons ici concerne la clause d'additivité des événements à 1, cependant il faut noter qu'il existe un Dutch book pour chaque clause probabiliste, voir Peterson (2009), p 157.

² Ramsey (1926).

³ Ibid. Nous trouvons aussi une théorie du pari chez Savage (1954), p 70-73. Le concept de pari sera la pierre angulaire des expériences de Lichtenstein et Slovic en 1971.

⁴ Ramsey, op. cit., p 175.

⁵ Voir, Savage (1954), p 82-91 ; Binmore (2009), p 116 et suivantes. Il faut remarquer que le concept de limitation, que nous allons aborder joue aussi dans un autre sens que nous étudierons plus loin, par le biais du concept de rationalité limitée, voir Simon (1956) ; Gigerenzer (2008) et (2011).

de l'utilité que nous pouvons leur attribuer¹ ; d'autre part, toutes nos décisions ne se situent pas sur le même plan : il y a toute une amplitude qui va des automatismes, en passant par les décisions habituelles jusqu'aux situations exceptionnelles. Et, dans cet entre-deux très vaste et très nuancé les lois de la cohérence et de la rationalité ne s'appliquent pas toutes de la même façon. C'est ce qu'avait vu Savage quand il expose comment s'appliquent ses axiomes dans des « mondes » très limités², et ce sera très important pour le reste de notre thèse que d'analyser l'applicabilité des principes du choix rationnel en fonction des situations qui se présentent aux sujets.

Les deux dernières parties de l'article de Ramsey, dans lesquelles nous n'entrerons pas de façon détaillée, sont destinées à justifier la valeur et l'usage de la logique inductive, et ce contre les vues contemporaines de Wittgenstein. Pour notre auteur, il y a, à la différence de Keynes, une distinction radicale entre les jugements inductifs et déductifs : les jugements inductifs sont synthétiques et augmentent notre connaissance, alors que les jugements déductifs sont analytiques et ne nous apprennent rien sur le monde³. Un argument inductif est un argument par lequel j'accrois mon savoir tout en étant en possibilité de le corriger :

« Nous devons donc expliquer comment, exactement, l'observation devrait modifier mes degrés de croyance ; bien sûr, si p est le degré observé, mon degré de croyance en q après l'observation devrait être égal à mon degré de croyance en q étant donné p avant ou, par la loi de multiplication, au quotient de mon degré de croyance en pq par mon degré de croyance en p »⁴.

Il s'agit ici d'appliquer le théorème de Bayes au problème de la révision des croyances⁵, Ramsey place ce procédé sous l'appellation générale d'une « logique humaine » qui doit constituer le complément nécessaire de la logique formelle classique. La logique inductive a pour vocation de répondre à la question : qu'est-il justifié de croire ? C'est une question pragmatique si nous faisons de la croyance le principe qui nous dispose en vue de

¹ Voir, Savage (1954), p 77 et 81.

² Voir, Savage, op. cit., p 84-91.

³ Auquel cas Ramsey rejoint ici Wittgenstein, pour qui la logique est tautologique et ne nous dit rien sur le monde. La distinction analytique/synthétique a ici une nette coloration kantienne, voir *Kritik der reinen Vernunft*. Einleitung, Seite 56-63.

⁴ Ramsey, op. cit., p 182.

⁵ Savage consacre quelques passages de son livre à ce problème, notamment p 45 et p 105-119 sur l'observation. En ce qui concerne les problèmes liés à la révision des croyances et leurs rapports à la conditionalisation, voir Lepage (2011).

l'action dans des circonstances définies.

L'induction est conçue par Ramsey comme un processus habituel : « L'induction est une de ces habitudes utiles, de sorte qu'il est raisonnable de l'adopter »¹. Il faut remarquer l'originalité de la démarche de l'auteur dans la défense de la logique inductive comme base pour la détermination de l'action, cette thèse sera relayée par Savage dans son maître ouvrage d'où les préoccupations sur la logique inductive ne sont pas absentes².

Nous concluons cette partie consacrée à Ramsey en allant dans deux directions. Tout d'abord une première question, préparatoire, émerge quand nous nous situons dans le cadre général de notre travail : après la critique des axiomatisations de la décision que vont mener les économistes et les psychologues à partir des années 50, que pourra-t-on sauver de la théorie de Ramsey, mais plus généralement que pourra-t-on sauver de la théorie de l'utilité en général ? Nous savons que les critiques vont axer leurs attaques sur l'axiome d'indépendance que Ramsey formule sans le formaliser. Nous pourrions postuler que ce qui va rester valide ce seront les axiomes d'ordonnancement et de transitivité, puisque nous avons vu et verrons encore que les sujets prennent de véritables risques en les violant. Pourtant, il semblerait que ces principes élémentaires de cohérence recevront aussi du domaine de l'expérience leurs limitations, voire leurs réfutations³.

Deuxième remarque, comme nous l'avons noté dans notre introduction générale : l'axiomatisation de la décision au XXe siècle s'opère de concert avec une disparition du concept de volonté de la sphère du décidable⁴. Or c'est une déconnexion assez paradoxale et problématique, car est-il possible de penser les concepts de préférences et de décisions en faisant abstraction de toute référence au volontaire ou à l'involontaire ? C'est ce qu'inaugure ici Ramsey, car on ne trouve dans son article aucune trace du concept de volonté. Parce qu'en plus de ne pas être définie ou utilisée, l'idée de volonté n'est même pas ramenée à un élément

¹ Ramsey, op. cit., p 187. Nous ne sommes pas très loin ici du concept d'heuristique tel qu'il sera élaboré par Kahneman/Tversky (1974).

² Savage, op. cit., p 46-50, et p 159.

³ À ce sujet, les travaux d'Edwards et dans sa lignée ceux de Lichtenstein et Slovic vont s'avérer déterminants.

⁴ Il y eut pourtant des philosophies de la volonté au XXe siècle, comme en témoigne par exemple Ricoeur (1950), ou d'un tout autre point de vue, Davidson (1970). Ce n'est pas de la sphère de la philosophie que la volonté disparaît.

du système axiomatique.

Cependant, les critiques qui vont se positionner dans les années 50 contre les axiomatiques, et qui se continueront dans les années 70 mettront au jour des processus et des structures plus profondes que l'axiomatisation, même dans son continuel raffinement depuis Ramsey¹, avaient tout simplement ignorés. Et, c'est précisément autour de ces structures, rendues visibles par la critique des axiomatiques, que nous étayerons nos hypothèses concernant la reconstruction originale du concept de volonté

¹ Pour une point de vue exhaustif sur la façon dont les théories de l'utilité ont essaimées depuis von Neumann/Morgenstern, voir Barberà/Hammond/Seidl (2004), p 687-837.

B- Décision et utilité chez von Neumann et Morgenstern

1) Le jeu comme paradigme décisionnel et le problème de l'utilité.

Quel est l'intérêt du jeu pour la philosophie ? Qu'est-ce que le jeu nous apprend sur la décision humaine¹ ? Le jeu comme pratique, mais aussi comme objet de réflexion, traverse la totalité de la culture et de la philosophie occidentale d'Héraclite jusqu'à nous². Aussi n'est-il pas étonnant de lire dans Leibniz qu'« On ne saurait dire combien les jeux renferment d'enseignements précieux pour l'art d'inventer »³.

Enfin, nous savons, grâce aux objets et aux restes retrouvés dans divers sites archéologiques, notamment égyptiens, que les jeux de hasard, comme les osselets ou les dés sont des pratiques très anciennes⁴.

En ce qui concerne le problème de la décision et son traitement scientifique dans le cadre de la Théorie des Jeux (TJ), il allait falloir attendre le livre de von Neumann et Morgenstern : *Theory of games and economic behavior*, publié en 1944⁵. C'est dans un contexte bien particulier que le concept de jeu va être étudié : le contexte économique. L'idée centrale et parfaitement originale des auteurs est de faire du jeu le paradigme du comportement décisionnel des sujets en contexte économique : « We shall first have to find in

¹ Remarquons que la plupart des ouvrages modernes portant sur la décision réservent une partie importante à la Théorie des Jeux (TJ, à partir de maintenant) ; voir, par exemple, Peterson (2009), p 212-262 ; Binmore (2009), p 25-34. Cependant, ceci n'est pas une règle, comme en témoignent par exemple les ouvrages de Chernoff (1959) et Gilboa (2009). Il conviendrait de s'interroger de façon plus approfondie sur les rapports entre théorie de la décision et théorie des jeux, la seconde est-elle seulement une partie de la première ? Pour quelques réflexions séminales à ce sujet on peut toujours se reporter au classique de Luce/Raiffa (1957) ; ainsi qu'à Myerson (1991) ; et Gintis (2009).

² Nous ne pouvons pas, de façon stricte, parler de TJ scientifiquement avant Borel (1921) et (1923), von Neumann (1928) mais surtout von Neumann/Morgenstern (N/M à partir de maintenant) (1944). Cependant, dès Cardan, Galilée, puis Pascal et Huygens le jeu devient un objet central de la réflexion scientifique. Voir, Hacking (2006), p 49-92 ; Gigerenzer et Al. (1989), p 1-13. D'un autre point de vue, tout aussi intéressant et fondamental pour la pensée occidentale, le jeu va être (re)constitué en objet philosophique de premier plan par le phénoménologue Eugen Fink, dans son livre de 1960 *Le jeu comme symbole du monde*. Ouvrage dans lequel il reprend, prolonge et approfondit certaines intuitions d'Héraclite, notamment issues de l'aphorisme 52, et du *Zarathoustra* de Nietzsche. La fameuse métaphore du lancer de dé chez Nietzsche trouvera un écho original chez Hacking (1990), p147-149.

³ Lettre à Jean Bernoulli, 29 janvier 1697 ; voir aussi, au même, 5 mars 1697, et, à Montmort, 17 janvier 1716.

⁴ Voir, David (1955) ; Hacking (1975).

⁵ L'ouvrage a connu de nombreuses éditions, la dernière en 2007. Nous utilisons la version de 1947 reprise dans l'édition du 60e anniversaire en 2007 aux Presses de l'université de Princeton, elle contient de nombreuses et précieuses postfaces. Mais, surtout, c'est dans cette version que se trouve l'axiomatisation de l'utilité, absente de l'édition originale. Sans exagérer, nous pouvons dire que ce livre est l'acte de naissance officiel des théories de l'utilité et de la décision modernes. D'autant plus que l'article de Ramsey était encore largement ignoré à ce moment-là. On trouve une très bonne recension du livre dans Edwards (1954) ainsi que dans les différentes postfaces ajoutées lors de la publication de l'édition du soixantième anniversaire. Par ailleurs Luce et Raiffa (1957) proposent une présentation claire et précise. La lecture de Poundstone (1992) est aussi très instructive tant sur le contenu de l'ouvrage, expliqué de façon très pédagogique, que sur la personnalité de von Neumann.

which way this theory of games can be brought into relationship with economic theory, and what these common elements are »¹. Il ne s'agit pas chez N/M de penser le jeu d'une façon philosophique, ni de s'intéresser à son importance ethno-sociologique – liée à la sociologie du temps libre et du loisir² – ou métaphysique³. Il ne s'agit pas non plus d'offrir, à la façon de Cardan, des « recettes », des astuces, ou des conseils pour mieux jouer, et pour mettre toutes les chances de son côté⁴. L'intérêt du jeu est de nous offrir ce que nous pourrions appeler un ensemble de situations décisionnelles tout à fait épurées, un espace décisionnel neutralisé du point de vue axiologique, déconnecté du monde réel, en somme un microcosme économique. *Le jeu est l'exercice pur de la décision stratégique placé sous l'égide du conflit d'intérêts* et dans certains cas de la coalition⁵. Le concept de « conflit d'intérêts » est directement lié au fait que chaque joueur, qui ne maîtrise qu'un nombre limité de variables⁶, veut maximiser sa fonction d'utilité⁷. Le jeu est donc lié aux concepts de stratégie ou de conflit d'intérêts⁸, c'est là la définition même du jeu : un conflit d'intérêts *réglé*. En effet, ce qui définit le jeu de façon première chez N/M ce sont ses règles : « The game is simply the totality of the rules which describes it »⁹. De la même façon que les sujets sont censés maximiser leur fonction d'utilité en jouant il en ira de même en contexte économique où les règles vont devenir les contraintes

¹ N/M (1947), p 1. Quelques lignes plus haut dans ce texte, donc au tout début de l'ouvrage, les auteurs relient de façon directe le problème économique à la question de la maximisation de l'utilité attendue (MUA). Nous allons voir que cette idée est soumise à de sérieuses contraintes dans le domaine de la décision humaine. L'idée de maximisation sera critiquée par Allais (1953) et Edwards (1954), nous poursuivrons cette critique en nous demandant dans quelle mesure les sujets, en contexte décisionnel, se *représentent* de façon cohérente ce qui est à maximiser.

² Voir, Lafortune (2004), et pour un point de vue plus général Dumezadié (1973).

³ Ce que fait Fink dans son livre aux pages 67 à 123. Fink ne mentionne d'ailleurs *jamais* la théorie de N/M, et l'oubli est important... Nous pourrions dire que le livre de Fink est le complément, au sens ensembliste du terme, de celui de N/M : il propose une philosophie du jeu.

⁴ Le seul jeu que N/M étudient un peu dans ce sens-là est une forme de poker très épuré. Voir, N/M, op. cit., p 186-219. La stratégie consiste à bluffer même avec une petite main, et ce au cas où l'adversaire aurait une main encore plus petite.

⁵ La TJ sera donc une étude mathématique du conflit d'intérêts entre un nombre n de joueurs ; à ce propos le sous-titre de l'ouvrage classique de Myerson est d'ailleurs : « Analysis of conflict ». La TJ est une étude des interactions possibles entre les acteurs qui nous éloigne en fait beaucoup de l'aspect ludique originel du jeu en lui-même. Peut-être que la dénomination « TJ » est ici un peu malheureuse. Voir, Poundstone (1992), p 39, Luce/Raiffa (1957), p 3-10. Pour une présentation plus générale de la notion de conflit, voir Schelling (1960).

⁶ Dans notre troisième partie, nous nous interrogerons, sur le nombre, très fluctuant et très important de facteurs que le système de référence décisionnel du sujet peut et doit prendre en compte dans une situation donnée.

⁷ N/M, op. cit., p 11.

⁸ On trouve dans la littérature classique, notamment militaire, de bons exemples d'anticipation ou d'application « naïve » de la TJ ; c'est le cas entre autres de *L'art de la guerre* de Sun Tzu (-512 AV JC), ou du livre de Clausewitz (1886). La littérature romanesque nous offre aussi des exemples intéressants avec par exemple *Le joueur* de Dostoïevski en 1867 ; Poe (1902) ; Hesse (1943) ; et plus récemment Lepage (2008).

⁹ N/M, op. cit., p 49. Il s'agit là d'une définition générale. Le jeu est composé de différents éléments formalisés par le calcul et la théorie des ensembles, par exemple les joueurs, les stratégies, les imputations, les mouvements, l'information. Tous ces concepts sont interreliés. Voir, Luce/Raiffa (1957), p 39-154 pour une introduction générale sur les jeux à deux joueurs ; voir aussi, Myerson (1991), p 37-88 ; et Gintis (2009), p 30-45.

extérieures ou les « lois » du marché. Le jeu nous offre donc un schème épuré de la décision économique, et pourquoi pas de la décision tout court ? Il y a un *isomorphisme* entre le contexte ludique et le contexte économique : « This theory of games of strategy is the proper instrument with which to develop a theory of economic behavior »¹.

Cependant l'application de la méthode mathématique à l'économie et au concept d'utilité ne va pas de soi², et ce pour de multiples raisons. Ce sera l'enjeu de la partie 1.2³ de présenter ces difficultés tout en tissant un lien très fort entre la théorie économique et la théorie physique, c'est aussi un moment programmatique pour le reste de l'ouvrage⁴. Nous n'allons pas analyser de façon exhaustive ces considérations méthodologiques, cependant quelques remarques s'imposent.

Tout d'abord, si l'on veut parvenir à formuler mathématiquement les problèmes économiques de façon aussi précise que les problèmes physiques, il faudra prendre garde que l'économie est une discipline récente⁵ et que nous ne connaissons les interactions économiques qu'imparfaitement⁶. D'autre part, nous ne bénéficions pas des milliers d'observations ordonnées et distribuées sur des siècles qui ont, par exemple, précédées la naissance de la mécanique rationnelle⁷. La base empirique est donc très mince. À ces éléments s'ajoute aussi une objection que nous trouvons déjà chez Ramsey : dans quelle mesure les phénomènes humains et sociaux, dans leurs irréductibles complexités, sont-ils mathématisables ?

Enfin, comme les problèmes économiques ne sont pas clairement formulés conceptuellement, mais de façon vague, l'application d'un formalisme mathématique – qui reste à penser – est très difficile. Il allait donc falloir partir de problèmes très locaux,

¹ N/M, op. cit., p 2.

² Contrairement à ce que certains économistes comme Jevons pouvaient penser concernant la mathématisation quasi « naturelle » de l'économie. Voir, Stigler (1950), I, p 317.

³ Elle couvre les pages 2 à 8.

⁴ Nous pouvons dire que ce qui était implicite chez Ramsey, ou très nuancé, ou incomplètement développé, à savoir le rapport fondationnel avec la physique devient ici explicite. Comme nous allons le voir bientôt, il en ira de même concernant le problème de la mesure de l'utilité que N/M vont complètement expliciter, ce qui n'était pas le cas chez le philosophe britannique. D'autre part, il y aurait beaucoup à dire sur les tentatives d'importations du paradigme de la théorie physique dans d'autres disciplines, notamment à partir des travaux du Cercle de Vienne et sa « vision scientifique du monde ». Il s'agit en fait du difficile problème de la fondation scientifique des sciences de l'homme ; à ce sujet voir Granger (1967).

⁵ Voir Foucault (1966), p 177-224 ; Larrère (1992), p 173-221.

⁶ N/M, op. cit., 2. Les auteurs font remarquer qu'à l'heure où ils parlent, il n'existe pas de « système économique universel ».

⁷ Op. cit., p 4. Le rôle de l'observation dans la constitution de la mécanique rationnelle, bien qu'important, ne doit pas nous faire oublier que cette discipline possède une partie fondationnelle a priori très importante. Voir, Duhem (1906), (1908) et (1913-1915) ; ainsi que, Koyré (1939), (1957), (1965).

précisément délimités et définis, comme ce fut le cas en physique avec le phénomène de la chute des corps, et : « We shall find it necessary to throw upon techniques of mathematics which have not been used heretofore in mathematical economics »¹.

La notion de jeu offre des problèmes suffisamment locaux et épurés pour servir de base, non pas empirique, mais conceptuelle à la constitution formelle du problème de la décision en contexte économique². La complexité et la diversité des échanges humains et des phénomènes économiques sont telles, qu'à l'instar de la physique, il faut avant tout délimiter le champ d'investigation scientifique et diviser les difficultés³. Le procédé sera le suivant : «...to obtain first utmost precision and mastery in a limited field, and then to proceed to another, somewhat wider one, and so on »⁴. Comme le notent les auteurs, il s'agit d'une étape préliminaire et heuristique, une phase de transition entre le non-mathématique et la formalisation parfaitement rigoureuse, qui à terme devrait permettre à la science constituée d'atteindre son but ultime : la prévision⁵.

Il va donc falloir définir les concepts fondamentaux de façon parfaitement claire et explicite. Ce sera le cas du concept d'utilité. Elle est un concept premier pour la simple raison qu'elle permet de répondre à cette question : pourquoi les joueurs jouent-ils ? Pourquoi les joueurs préfèrent-ils une décision stratégique au détriment d'une ou de plusieurs autres ? La réponse est claire : pour maximiser leur utilité attendue. De la même façon que le concept de chaleur a été formalisé en partant d'une sensation originaire : « plus chaud que », se pourrait-il que l'utilité soit dérivable et formalisable à partir de la sensation originaire de « préférence »⁶ ?

Cette entreprise originale de formalisation allait du même coup devoir passer par une nouvelle définition du comportement rationnel en contexte de libre compétition. Dans le cadre de la théorie de N/M un sujet rationnel devra choisir entre des alternatives de façon à

¹ N/M, op. cit., p 5.

² Comme le note les auteurs page 5 : « the aim of the book lies not in the direction of empirical research ».

³ Nous retrouvons ici de façon explicite un des quatre préceptes cartésiens de la méthode : « To divide the difficulties (...), and to reduce all others as far as reasonably possible », comparer avec Descartes (1637), p 88-90, Ed. Livre de Poche.

⁴ N/M, op. cit., p 7. Il s'agit en quelque sorte de travailler par élargissements concentriques à partir d'un noyau dur parfaitement structuré, maîtrisé et formalisé. Dans le cours de l'ouvrage il s'agira de complexifier la théorie en fonction, par exemple, du nombre de joueurs et des enjeux, pour aller vers la formalisation du concept opératoire central de coalition.

⁵ N/M, op. cit., p 8.

⁶ N/M, op. cit., p16. Dans quelle mesure la relation de préférence peut-elle être considérée comme une sensation ? La question reste largement ouverte car il semblerait que nous soyons ici plus dans le registre de la réflexivité que de l'immédiateté.

maximiser une fonction d'utilité¹, ou sa fonction de profit s'il s'agit d'une entreprise : « The individual into attempts to obtain the respective maxima is also said to act 'rationally' »².

Afin de s'assurer que les sujets agissent comme si ils maximisaient leur fonction d'utilité, les auteurs vont proposer une axiomatisation de ce concept, elle sera beaucoup plus structurée et explicite que celle de Ramsey. Par ailleurs, ici, à la différence du philosophe britannique nous n'allons pas des différences de valeurs entre les options aux probabilités mais inversement : ce sont les probabilités, conçues cette fois de façon objective, qui détermineront les différences de valeurs entre les options. Cependant, cette axiomatique va être, comme chez Ramsey, pensée sur le modèle d'une théorie de la mesure³. Les sujets, s'ils satisfont les axiomes, choisissent comme s'ils maximisaient leur fonction d'utilité⁴. La rationalité sera donc définie comme chez Ramsey, en fonction du respect d'un certain nombre de clauses formelles par le décideur. Aussi convient-il de se pencher immédiatement sur cette reconstruction du concept d'utilité⁵, car elle constitue, avec les théories de Ramsey et Savage, un moment central et nécessaire pour comprendre sur quels éléments vont porter les critiques de Kahneman et Tversky.

2) Une nouvelle conception de l'utilité et de la décision.

Comme nous l'avons vu dans l'introduction de ce chapitre, c'est à proprement à partir de Bernoulli (1738)⁶ que commence la TUA, et nous pouvons dire que, hormis quelques

¹ Nous verrons dans notre deuxième partie, consacrée à la critique de la TUA, que c'est en particulier à propos de ce concept – qui a une importante connotation normative – que les désaccords vont émerger, et ce très rapidement. Friedman/Savage (1948) signalent que dès la parution de leur livre, N/M furent confrontés à des problèmes importants, que tente d'ailleurs de résoudre l'article de Friedman/Savage. Ces problèmes sont notamment liés à deux phénomènes courants dans le comportement des sujets : l'achat simultané d'assurances de diverses sortes et de tickets de loteries. Or ces deux faits sont incompatibles avec l'hypothèse de MUA. Dans la Prospect Theory de 1979, Kahneman et Tversky reviendront abondamment sur ce type d'incompatibilité entre la TUA et le comportement réel des sujets. Le concept de comportement rationnel sera le centre de la plupart des critiques qui iront d'Allais (1953) jusqu'à Slovic/Lichtenstein (1968), et Kahneman et Tversky (1974), (1979). Les problèmes liés à la rationalité et à ses limites seront aussi étudiés, dans une toute autre direction – celle de l'adaptativité – par Simon (1956) et ses successeurs comme Gigerenzer (2011).

² N/M, op. cit., p 9.

³ N/M, op. cit., p 20-21, section 3.4.

⁴ Voir, Picavet (1996), p159-162 ; Binmore (2009), p 39-60 ; Gilboa (2009), p 78-86 ; Peterson (2009), p 91-106. Ce sera le réquisit central de toutes les théories de l'utilité qui succéderont à N/M, y compris celle de Savage (1954). Pour une présentation précise des théories qui ont été proposées à la suite des critiques adressées à N/M, voir R. Sugden in Seidl/Hammond/Barberà (2006), p 687-755.

⁵ Voir, Friedman/Savage (1948), p 281 ; Friedman/Savage (1952) ; et, Savage (1954), p 97 et suivantes.

⁶ Sur l'apport de cet auteur on peut se reporter à Savage (1954), p 91-96 ; Daston (1988), p 70-77 ; Gigerenzer et Al. (1989), p 12-18 ; Picavet (1996), p 152-158.

remarques de Laplace (1886) concernant le concept d'espérance morale¹, les hypothèses de Bernoulli resteront lettre morte jusqu'à la parution du livre de N/M².

Il va donc s'agir de reconstruire l'hypothèse de maximisation de l'utilité attendue dans le cadre d'un formalisme mathématique nouveau dont les deux concepts centraux seront d'une part les probabilités conçues objectivement et d'autre part, l'ensemble³ des conséquences, des rétributions, ou des prix. Ce qui correspond au concept de monde possible chez Ramsey. Comme nous le verrons plus tard, c'est par l'institution d'une relation entre les probabilités et les conséquences, matérialisée par le concept de loterie, et dans l'axiomatisation de l'utilité, que réside l'innovation majeure de N/M. Comme le remarquent les auteurs, l'utilité ne peut donc être simplement et immédiatement ramenée à une sensation originaire qui permettrait de la mesurer. La « sensation » de préférence peut seulement nous fournir une indication ordinale qui bien qu'importante du point de vue introspectif n'a pas grand intérêt du point de vue scientifique : « It is not itself a basis for numerical comparison of utilities for one person nor of any comparison between different persons »⁴.

À l'instar de Ramsey, nous pouvons dire que l'introspection ne nous est ici d'aucun recours. Les problèmes sont en quelque sorte semblables à ce qui s'est passé avec le phénomène de la chaleur dans sa phase préthéorique, il était impossible de dériver quoi que ce soit de numérique à partir de la sensation originaire : un corps est plus chaud qu'un autre⁵. Pourtant l'étude approfondie de ce phénomène a permis par la suite de se diriger dans deux directions : quantité de chaleur et échelle des températures. Pourquoi ne pourrait-il pas en aller de même avec l'utilité ? Que faudrait-il ajouter à la sensation originaire de préférence pour

¹ Concernant l'histoire du concept d'utilité, on peut se reporter à l'article classique et très exhaustif de Stigler (1950), en deux parties. Cet article part de Bentham – et de sa critique par les ricardiens – ainsi que de Bernoulli et retrace toutes les étapes de la constitution du concept de l'utilité et des problèmes liés à sa mesure dans le domaine de l'économie jusqu'à Marshall. Savage (1954), consacre une analyse intéressante à l'histoire de la TU aux pages 91 à 103 de son ouvrage, bien entendu dans le cadre de Savage il s'agit de l'utilité au sens bernoullien du terme, c'est-à-dire avec l'hypothèse de maximisation. Voir aussi Allais (1953), p 506-514.

² Stigler mentionne et analyse avec raison l'influence indéniable de la théorie de Bernoulli (1738) sur les recherches psychométriques de Weber et Fechner, notamment concernant l'utilisation de la fonction logarithmique pour représenter l'utilité chez le premier, et la sensation chez les suivants. Voir Stigler (1950), II, p 375 et Fechner (1889), chapitre XIII. Nous reviendrons sur la thèse de Weber/Fechner plus loin dans notre travail, lorsque nous étudierons le concept de seuil psychologique en contexte décisionnel.

³ Il est crucial que nous ayons à faire à un ensemble pour les rétributions et à une mixture quand interviennent les probabilités, un ensemble au sens mathématique strict, avec la relation fondamentale \in . En effet, comme le remarque Granger (1988), p 256 et suivantes, seuls des éléments qui font partie d'un ensemble au sens strict peuvent faire l'objet d'une mesure mathématique.

⁴ N/M, op cit, p 16. Il y a un aspect machien – dû à l'idée de rapprocher les théories de sensations originaires au niveau fondationnel – dans la façon dont les auteurs conçoivent la constitution de leur théorie et de la théorie physique.

⁵ Voir Carnap (1960), p 62-69.

passer du qualitatif au quantitatif ? Pourquoi pas un coefficient de probabilité et une échelle d'utilité d'un genre particulier qui soit à la fois ordinale et cardinale ¹?

N/M vont maintenant commencer à s'intéresser au problème de l'introduction des probabilités dans la formalisation du processus de décision et à la combinaison des événements entre eux en fonction des probabilités. Si nous prenons B et C pour symboliser les événements et la distribution de probabilités ($1/2$; $1/2$) : « Then the 'combination' is the prospect of seeing B occur with a probability of 50 % and (if B does not occur) C with the (remaining) probability of 50 % »².

Dans chaque situation décisionnelle, comme les événements sont exclusifs, la probabilité de deux événements disjoints doit être égale à la probabilité de leur somme. Enfin, comme chez Ramsey, le décideur doit être capable d'ordonner toutes les options possibles, conçues sur le mode de tickets de loterie³.

Avec ces quelques éléments nous avons déjà de quoi comparer, à un niveau certes très élémentaire, des préférences entre elles. Imaginons un sujet qui soit confronté à la perspective suivante : A sûr, ou bien la loterie comprenant comme éléments ($1/2, B$; $1/2, C$)⁴. Si le sujet préfère A à B et B à C alors il préférera A à la loterie. D'autre part, si il préfère B à A et est indifférent à C , alors il choisira la loterie. Enfin, que se passe-t-il si ses préférences sont du type : A est préféré à B et C est préféré à A . « Then any assertion about his preferences of A against the combination contains fundamentally new information »⁵. En effet, information qui est la suivante : la préférence du sujet de A sur B est *plus grande* que celle de C sur A . Ce qui signifie que les différences entre les options deviennent mesurables. Nous sommes très proches de Ramsey lorsqu'il tentait de modéliser la mesure des différences de valeurs entre les

¹ L'exposé que nous étudions ici, allant de la page 16 à la page 24, et qui est consacré au problème de la fondation d'une théorie de la mesure de l'utilité, est capital. Cependant, il n'a pas été, à notre connaissance, étudié de façon minutieuse et systématique. La plupart des livres et des manuels tiennent cette phase préparatoire pour acquise, ou ne la mentionnent même pas. Du point de vue philosophique, le seul à étudier le problème est Granger (1967), p 137-138, (1988), p 255-261, qui s'appuie sur des éléments de théorie de la mesure que l'on trouve chez Suppes/Zinnes (1963), p 2-76. Même des ouvrages aussi pointus que Fishburn (1970), (1973), ou Kreps (1988), et des articles aussi précis que Friedman/Savage (1948), (1952), Marschak (1950), Samuelson (1952), Herstein/Milnor (1953) ne comportent aucune remarque ni discussion approfondie à propos de cette partie fondamentale et fondatrice. Or, c'est seulement en tant que la théorie de l'utilité est une théorie de la mesure qu'elle a vraiment un sens et une pertinence. N'oublions pas la thèse de Ramsey (1926), qui vaut encore ici : c'est la mesure qui donne sens à ce qui est mesuré, et cette thèse restera valide dans le cadre de Savage (1954), p 69-91.

² N/M, op. cit., p 18.

³ Binmore (2009) propose une explication très pédagogique de ce concept aux pages 39 et 43. L'artifice graphique qu'il utilise est très intéressant d'autant plus qu'il permet de « visualiser les axiomes ».

⁴ Perspective qui d'un point de vue plus général a la forme suivante : $[A \vee (Bp ; C(1 - p))]$. Luce/Raiffa (1957), analysent un exemple similaire aux pages 21-22.

⁵ N/M, op. cit., p 18.

options, et ce même si le philosophe britannique utilisait un vocabulaire assez différent, par exemple le concept de monde possible représenté ici par les mélanges d'événements et de probabilités. Pour N/M, il s'agit d'une première étape dans la quantification de la relation de préférence. La mesure probabiliste sera représentée par un nombre réel α compris dans l'intervalle $[0,1]$ ce nombre sera *distribué* entre les événements d'une même option et additif à un, ainsi : « We suggest the use of α as a numerical estimate for the ratio of the preference of A over B to that C over B ».

N/M vont maintenant aborder une autre question : du point de vue temporel les choix des sujets doivent être localisés dans un futur immédiat, et ce afin d'éviter les problèmes liés à l'impact du concept de délai sur les préférences. C'est là, la première limitation philosophique de la théorie de N/M. En effet, elle ne prend pas en compte le temps – c'est une théorie statique – ni l'aspect dynamique des processus décisionnels. Or, comme nous le montrerons dans notre dernière partie, le temps, ainsi que le concept de délai sont des facteurs centraux, voire déterminants, de toute prise de décision. Ces éléments seront repensés en fonction du problème général de la *représentation* et des phénomènes plus localisés de pondération¹.

À la suite de cette réflexion sur les rapports des choix rationnels au temps, N/M vont passer à un autre problème, tout aussi central dans le domaine de la théorie de la décision : l'interprétation des probabilités². Et, ce n'est pas un hasard si c'est l'interprétation fréquentiste³ qui est retenue : « The simplest procedure is, therefore, to insist upon the

¹ Il est bien évident que les préférences évoluent en fonction du temps. À cet égard, l'exemple qu'analyse Picavet (1996), p 167 est intéressant. L'idée étant que dès que nous nous éloignons du futur proche, nos préférences peuvent alors ne plus être cohérentes du point de vue de la TUA et même violer certains axiomes comme l'axiome d'indépendance par exemple. Le rapport du temps avec cet axiome est aussi analysé par Samuelson (1952), p 674-677. Les relations entre temporalité et décision ont été étudiées, dans le cadre de la TUA par McClennen (1990). Thaler (1981) étudie ces phénomènes du point de vue de l'inconsistance dynamique ; voir aussi Read (2004), en particulier concernant le rapport du temps avec la fonction de valeur ; enfin on peut se reporter aux deux articles séminaux de Loewenstein/Prelec (1992) et (1993) qui eux se situent dans le cadre de la Prospect Theory, et dont Read s'inspire par ailleurs. Dans la suite de notre étude nous voudrions montrer qu'il existe un *coefficient de pondération temporel* et quels sont les effets de ce coefficient sur la prise de décision.

² Nous ne rentrerons pas dans le débat sur l'interprétation des probabilités, il faut seulement noter l'émergence simultanée des points de vue subjectifs et objectifs aux XVIIe et XVIIIe siècles ; voir Hacking (2006), p 11-17, et (2004), p 140-161 ; et, Daston (1988), p 210-225. Il est intéressant de remarquer, que pour Hacking, les deux points de vue émergent de façon simultanée, alors que pour Daston la séparation, bien que virtuelle dès les origines, est consommée à partir de Condorcet et Laplace. Ce que nous pouvons ajouter ici c'est que les tensions entre « subjectivistes » et « objectivistes » sont toujours très présentes dans le débat philosophique contemporain ; en particulier depuis l'essor majeur qu'a connu le Bayésianisme avec Savage (1954) et Jeffrey (1983) notamment dans l'application des processus de conditionalisation à des problèmes épistémiques et doxologiques généraux. Savage réserve d'ailleurs les chapitres 6 à 17 de son ouvrage à ces questions. Cependant, selon notre point de vue et en suivant, entre autres, Hacking (1971), Gillies (2000) et Binmore (2009), les deux interprétations sont complémentaires et ne s'appliquent pas aux mêmes domaines d'investigation. Pour un point de vue complet sur les avancées du Bayésianisme ; voir, Oaksford/Chater (2006) et Chater/Oaksford (Eds.) (2008).

³ Voir, Gillies (2000), p 88-112 ; Hacking (2004), p 205-265. Rappelons simplement que dans cette interprétation ce qui nous intéresse c'est la *fréquence* de l'apparition d'un phénomène dans un ensemble de référence donné. Par exemple, le nombre de

alternative, perfectly well founded interpretation of probability as frequency in long runs »¹. Il est possible que cette alternative théorique ait été choisie du fait de l'application que von Neumann avait fait de l'interprétation fréquentiste dans le domaine de la physique quantique, dans son ouvrage de 1932 *Mathematical foundations of quantum mechanics*. Nous verrions donc ici, le parallèle avec la théorie physique se poursuivre et s'accroître.

Le problème de l'interprétation des probabilités en théorie de la décision nous amène au problème général du contexte décisionnel. Du point de vue de l'économie il avait été défini par Knight en 1921², il désigne l'état de connaissance des sujets concernant les probabilités dans une situation de choix. En contexte d'*incertitude*, les probabilités sont données de façon objective et l'événement « arrive » accompagné d'une probabilité de réalisation donnée, c'est-à-dire sur le mode d'un billet de loterie : $(\alpha, p; \beta, 1-p)$. Dans le cas de l'*ignorance* c'est au sujet lui-même de proposer une évaluation probabiliste subjective de l'événement. Le contexte d'incertitude correspond à l'axiomatique de N/M, celui d'ignorance est relié aux théories de Ramsey et de Savage³.

À la suite des remarques préparatoires que nous avons évoquées et un peu prolongées, les auteurs vont se tourner vers l'axiomatisation de l'utilité. Car c'est bien là que se situe le centre de gravité de leurs réflexions sur ce concept. Mais, avant cela, N/M reviennent une dernière fois sur le problème de la mesure, en l'affinant. Le parallèle avec la physique va donc être encore poussé un peu plus loin. En fin de compte, c'est la forme de la rationalité elle-même qui se trouve définie en parallèle avec la théorie physique. Le cheminement est le suivant : la physique constitue le modèle théorique de l'axiomatisation de l'utilité, et la rationalité est définie en fonction du respect des axiomes par le sujet.

L'axiomatisation, du fait de son lien essentiel avec la physique, va garantir, outre l'unicité de la fonction d'utilité à une transformation linéaire près, la *mesurabilité* de l'utilité.

pièces ou de faces dans une *séquence* de lancers de pièce précis. Émergent ici, de façon centrale et problématique les concepts de répétition et de long terme dont l'utilisation en théorie de la décision pose de nombreux problèmes.

¹ N/M, op. cit., p19.

² Voir, Ellsberg (1961) ; Picavet (1996), p 170-171 ; Peterson (2009), p 5.

³ La distinction classique entre les différents contextes va commencer à être remise en cause par Ellsberg (1961). On trouve d'excellentes présentations des différents contextes décisionnels dans Luce/Raiffa (1957), p 12-23 et p 275-326 pour le contexte d'incertitude (on parle aussi de contexte d'ignorance). L'approche de Peterson, très pédagogique, est intéressante car elle permet de comprendre comment, dans certains cas bien définis nous pouvons passer du contexte d'ignorance au contexte de risque en appliquant de principe de raison insuffisante (PRI), par exemple, p 53 et suivantes. Voir aussi, Hacking (2004), p 156.

Mesurabilité qui pourra être considérée comme parfaitement objective, en partie du fait de l'interprétation des probabilités choisie par les auteurs. Autrement dit : *tout sujet rationnel face à une alternative donnée devra décider de façon à maximiser son utilité attendue*. Ce qui implique que des sujets différents dans une situation identique devront opter pour les mêmes options. C'est là une autre limitation majeure de cette théorie, aussi bien du point de vue descriptif que normatif¹, car elle ne prend pas en compte la singularité de la fonction de valeur et du système de référence décisionnel des sujets.

3) L'axiomatisation de l'utilité.

Il s'agit là du moment central et séminal de la théorie moderne de la décision en contexte de risque, séminal en ce sens que c'est à partir de cette axiomatique que vont se situer toutes les autres théories qui entendront formaliser le concept d'utilité, et ce dans des domaines aussi variés que l'économie, la psychologie ou la philosophie². La TUA subira des modifications assez importantes du fait de sa confrontation avec le domaine empirique, c'est-à-dire avec le comportement des sujets soumis à des alternatives réelles³.

Dans le domaine de la philosophie de langue anglaise, c'est une théorie qui aura une influence sur Davidson⁴, Jeffrey⁵, Rawls (1971), Nozick et Lewis. Nozick étudiera l'opposition entre la TUA et le concept de dominance dans le cadre du problème de Newcomb⁶, tandis que Lewis proposera ses hypothèses dans le cadre d'une théorie causale de la décision⁷. Dans la philosophie de langue française, les contributions les plus significatives à l'étude de la TUA sont Granger (1967) et (1988), Saint-Sernin (1973), et enfin la thèse d'Emmanuel Picavet (1996).

Ce qui va nous intéresser ici, à l'intérieur de notre recherche, et qui nous intéressait

¹ Sur ces concepts, voir Baron (2009), p 31-34.

² En ce qui concerne les nombreuses théories alternatives qui ont été proposées depuis N/M (1944) on peut se reporter aux chapitres 14 et 15 de Barberà/Hammond/Seidl (2004), p 687-837. Dans le chapitre 14, R. Sugden analyse l'aspect fondationnel en relation avec les contestations empiriques ; dans le chapitre 15, U. Schmidt s'intéresse lui à l'aspect formel et aux aménagements mathématiques qu'il a fallu entreprendre au fur et à mesure que les alternatives à la TUA classique ont émergées.

³ Voir, Saint-Sernin (1973), p 75 ; Picavet (1996), p 229-266.

⁴ Davidson a écrit un article sur la théorie de l'utilité avec Suppes en 1956, sur le rapport entre l'utilité et les probabilités subjectives. Certains essais de Davidson (1980) traitent aussi de la théorie de la décision et de la théorie de l'utilité.

⁵ Jeffrey utilise le modèle N/M dans son ouvrage de 1965, p 44-46.

⁶ Voir, Nozick (1969).

⁷ Comme en témoigne entre autres l'article édité par Gärdenfors/Sahlin (1988), p 377-405 : « Causal Decision Theory ».

déjà chez Ramsey, est un peu différent de ce que les philosophes sus cités ont étudié jusqu'à maintenant. Nous pouvons dire, sans mauvais jeu de mots, que ce qui nous intéresse ici c'est *l'utilité négative* de la TUA. En effet, dans le cadre de notre thèse, la TUA constitue le centre de gravité en ce sens que c'est seulement grâce à sa critique par l'économie et la psychologie cognitive que seront mis en avant les éléments subjectifs tels que l'aversion au risque ou les pondérations probabilistes sur lesquelles nous allons fonder notre conception du Système de Référence Décisionnel.

N/M vont maintenant passer à une « discussion plus détaillée des principes de la mesure »¹. C'est une partie qui a pour vocation de répondre à la question suivante : sur quel type de modèle issu de la physique pourrait être formalisé le concept de loterie² ? Nous savons déjà qu'il s'agit de *mélanger* des événements appartenant à un ensemble – au sens mathématique du terme – à des distributions de probabilités comprises dans l'intervalle [0,1]. Pour formaliser le concept de loterie il va donc falloir pousser la comparaison avec la physique un peu plus loin.

C'est à partir de la mesure du centre de gravité de deux corps dans un repère euclidien que les auteurs vont élaborer, de façon analogique, leur concept de loterie. Ce seront donc les notions de position et de masse qui vont servir d'*analogon* aux concepts d'événements et de distribution de probabilités :

« 'Position' as well as the vectorial quantities became triplets of numbers, called coordinates or components respectively. The 'natural' concept of 'center of gravity' of two positions $\{x_1, x_2, x_3\}$ and $\{x'_1, x'_2, x'_3\}$, with 'masses' $\alpha, 1-\alpha$, becomes $\{\alpha x_1 + (1-\alpha)x'_1, \alpha x_2 + (1-\alpha)x'_2, \alpha x_3 + (1-\alpha)x'_3\}$ »³.

Nous reconnaissons ici la structure formelle de ce qui deviendra une loterie avec comme composants les événements et les probabilités. Soit $\alpha, 1-\alpha$ une distribution de

¹ N/M, op. cit., p 20.

² Les remarques de Granger (1988) sont à ce sujet très éclairantes : si dans la physique toute tentative de mesure peut être ramenée de façon analogique aux concepts de longueur, de temps ou de masse, « Il n'est pas de même dans le domaine des sciences de l'homme, où l'application directe du système opératoire soulève immédiatement des objections », p 255. C'est précisément l'objection que N/M se sont adressés à eux-mêmes. Il va donc falloir dépasser cet obstacle pour trouver une méthode originale de mesure dont la proximité structurelle avec la physique devrait garantir la validité.

³ N/M, op. cit., p 21.

probabilités prise dans l'intervalle fermé $[0,1]$, et u,v des éléments de l'ensemble U des conséquences ou des prix. Une loterie L aura la forme :

$$[\alpha u, (1-\alpha)v]^1$$

Il faut garder à l'esprit que ce qui motive le choix de cette structure est son aspect naturel : «These 'natural' relations are the best base upon which to construct mathematical models and to correlate the physical domain with them»². Il serait toutefois intéressant de questionner plus profondément cette « naturalité » des relations physique qui semble allait presque de soi pour N/M.

Enfin, il conviendra de s'interroger sur le problème suivant : la corrélation des quantités physiques – ou des utilités ³ – avec des nombres est-elle unique ? Si ce n'est pas le cas il faudra s'assurer que les différentes façons de corréler les quantités mesurées avec les nombres soient équivalentes entre elles, c'est-à-dire que l'on puisse les *transformer* les unes dans les autres⁴ ; un peu à la façon dont, par exemple, on passe de l'échelle Fahrenheit à l'échelle Celsius⁵. Le but étant, dans le domaine de l'utilité, de préserver la relation de préférence ainsi que la mesure de la différence entre les options.

Une dernière étape avant l'axiomatique, liée aux problèmes d'unicité précédemment évoqués, consiste à s'assurer qu'on peut établir :

« a correspondance between utilities and numbers which carries the relation $u > v$ and the operation $\alpha u + (1-\alpha)v$ for utilities into the synonymous concept for numbers »⁶.

¹ La seule chose qui change ici est que les événements sont repérés par une seule lettre, alors que les positions dans la formule précédente étaient repérées par trois lettres du fait qu'il s'agit d'un système de référence à trois axes. Remarquons aussi qu'à partir d'une loterie simple nous pouvons construire n'importe quelle loterie composée et inversement, du moment que les axiomes du calcul des probabilités soient respectés, par exemple : $\{\beta[\alpha u, (1-\alpha)v], [1-\beta]w\}$; voir Luce/Raiffa (1957), p 26.

² N/M, *ibid*.

³ « For utilities the situation seems to be of a similar nature », N/M, *op. cit.*, p 23.

⁴ Encore une fois nous pouvons nous référer à Granger (1988) qui résume magistralement le problème : « La condition fondamentale de quantification étant l'existence d'une application de l'ensemble phénoménal dans un ensemble de nombres, l'une des questions qui se présentent tout d'abord est celle de son *unicité*. D'une manière plus précise, le morphisme choisi comme définissant tel type de quantification subsiste-t-il lorsqu'on le compose avec certaines transformations de l'ensemble numérique en lui-même, ou au contraire est-il unique, et lié à une application bien déterminée des objets vers les nombres ? », p 257. C'est précisément la question que N/M se pose dans la partie que nous analysons.

⁵ On sait que l'on peut passer d'une échelle à l'autre en utilisant la transformation affine suivante : $f = 9/5c + 32$. Dans le cadre de l'utilité, l'analogie avec les échelles de température fonctionne assez bien.

⁶ N/M, *op. cit.*, p 24. Voir aussi la citation de Granger de la note 2 page 60.

Il s'agit de montrer qu'il existe une correspondance entre l'utilité u et un nombre réel, correspondance qui produit un autre nombre : $V(u)$ qui est en fait l'utilité attendue de u . Toute la démonstration qui suit est importante : c'est la définition mathématique de la fonction d'utilité. Sa nature sera telle *si et seulement si* elle satisfait les axiomes. En résumé, formellement, le cheminement est le suivant :

$$u \rightarrow \rho = V(u)$$

Cette correspondance doit satisfaire deux conditions essentielles¹ : tout d'abord elle doit préserver l'ordre, c'est-à-dire que :

$$1) (u > v) \equiv (V(u) > V(v))$$

D'autre part, V doit être linéaire par rapport aux probabilités, c'est-à-dire varier en fonction des probabilités, ce qui veut dire avoir la propriété d'espérance :

$$2) V(\alpha u + (1-\alpha)v) = \alpha V(u) + (1-\alpha)V(v)$$

À partir de là, les auteurs postulent que deux correspondances existent :

$$u \rightarrow \rho = V(u)$$

$$u \rightarrow \rho' = V'(u)$$

En utilisant ρ et ρ' on montre ensuite que ρ' peut être exprimé en fonction de ρ . N/M montrent donc que :

$$\rho' = \varphi(\rho)$$

Ce faisant, φ doit respecter les propriétés 1) et 2) pour tout ρ et pour tout σ , c'est-à-dire pour tout nombre, et pour toute distribution de probabilité α , $(1-\alpha)$:

¹ Voir, Fishburn (1970), p 78 ; Kreps (1988), p 43.

$$1') \rho > \sigma \equiv \varphi(\rho) > \varphi(\sigma)$$

$$2') \varphi(\alpha\rho + (1-\alpha)\sigma) = \alpha\varphi(\rho) + (1-\alpha)\varphi(\sigma)$$

$\varphi(\rho)$ doit donc être une fonction linéaire. Nous limitons ici le raisonnement de N/M à ses étapes essentielles. Concernant l'utilité le cheminement sera le suivant : si la fonction d'utilité respecte les propriétés 1) et 2) – ce sera aux axiomes de le garantir – alors : « Utility is a number up to a linear transformation »¹. Il est donc crucial que pour garantir 1) et 2) un certain nombre de clauses soient introduites². Ils sont divisés en trois groupes comme chez Ramsey³ et presque tous expriment et formalisent des propriétés que nous avons rencontrées chez le philosophe britannique, mais à cette différence que les axiomes de N/M sont beaucoup plus explicites et beaucoup plus élaborés formellement⁴.

Passons à leur analyse : « We consider a system U of entities $u, v, w \dots$ In U a relation is given, $u > v$, and for any number α , ($0 < \alpha < 1$), an operation :

$$\alpha u + (1 - \alpha v) = w \gg^5.$$

Une loterie, une alternative, est représentée par le mélange entre les éléments de l'ensemble U avec une mesure probabiliste située dans l'intervalle fermé $[0,1]$. L'ensemble U

¹ N/M, op. cit., p 24.

² Voir, Saint-Sernin (1973), p 76 Picavet (1996), p 159. En fait la relation entre 1), 2) et les axiomes va dans les deux sens : les axiomes impliquent l'existence d'une fonction d'utilité caractérisée par 1) et 2) et réciproquement. Il existe plusieurs démonstrations de cette double implication, à notre sens la plus claire et la plus pédagogique est celle de Peterson (2009), p 302-307.

³ Les axiomes sont présentés aux pages 26 et 27 de l'édition de 1947. Nous allons préserver la notation originale pour la présentation, mais nous utiliserons une notation plus moderne par la suite.

⁴ À ce jour on peut trouver d'excellentes analyses de l'axiomatique N/M dans Friedman/Savage (1948) et (1952) ; Fishburn (1970), p 103-128 ; Kreps (1988), p 43-70. Ces deux derniers ouvrages se recoupent en bien des points, et Kreps se réfère souvent à Fishburn, en particulier dans la partie sur Savage. Les deux ouvrages doivent beaucoup à l'article séminal de Herstein/Milnor (1953) en ce qui concerne les mixtures (mixture sets). Du côté français, Granger (1955) contient des indications intéressantes qui seront reprises et approfondies dans certaines parties de ses ouvrages de 1967 et 1988. Il y a aussi Saint-Sernin (1973) et Picavet (1996) qui sont certes clairs, mais un peu courts sur l'axiomatique N/M. Cependant, Picavet propose une démonstration originale de l'hypothèse de maximisation à partir de l'axiome d'indépendance. Plus récemment en langue anglaise, il y a la présentation très claire et synthétique de R. Sugden dans Barbera/Hammond/Seidl (2006), p 689-694, cette présentation est elle aussi basée sur Herstein/Milnor (1953). Enfin, pour une approche plus pédagogique, on peut se reporter à Peterson (2009) ; Binmore (2009) ; et, Gilboa (2009) qui est un peu plus formel que les deux autres. Cependant ces quelques références qui nous semblent importantes sont loin d'épuiser le sujet et les recherches formelles dans le paradigme de la TUA continuent de progresser et d'intégrer de plus en plus d'éléments.

⁵ N/M, op. cit., p 26.

représente les utilités, les conséquences, les rétributions ou les prix¹. La loterie ainsi obtenue à partir du mélange de ces deux éléments est aussi un élément de U . Une perspective est le mélange d'une utilité u avec une distribution de probabilités : $\alpha, (1-\alpha)$. C'est en fait *l'injection* d'une probabilité objective qui rend l'utilité *mesurable*. Ce qui soulève une question philosophique importante : dans quelle mesure le probable peut-il être considéré comme la mesure du possible² ? Une loterie est donc une mixture³, c'est-à-dire un ensemble non vide qui associe à toute mesure probabiliste sur $[0,1]$ et à toute paire u, v , appartenant au produit cartésien de U par lui-même, un élément $\alpha u + (1 - \alpha)v$ ⁴. La valeur de la loterie est représentée par « w » qui est une classe d'équivalence⁵ pour tout événement pouvant être obtenu par la combinaison : $\alpha u + (1 - \alpha)v$ ⁶. Une loterie peut donc être identifiée à une de ses classes d'équivalence et réciproquement.

La deuxième étape de l'axiomatique consiste à montrer que la relation $>$ sur U et l'opération : $\alpha u + (1 - \alpha)v$ doivent respecter les axiomes suivants :

a – Ordonnement.

« (3 : A) $u > v$ is a *complete ordering* of U .

This means : write $u < v$ when $v > u$. Then

(3 : A : a) For any two u, v one and only one of the three following relations holds :

$u = v, u > v, u < v$ ⁷.

¹ Le fait de savoir si U est fini ou infini, dénombrable ou non ne sera pas abordé ici, mais c'est néanmoins un point important. Pour U infini, voir Herstein/Milnor (1953) ; et Kreps (1988), p 32-33. À la différence de Marschak (1950), Herstein/Milnor seront les premiers à considérer la possibilité que U soit infini, les auteurs parlent alors « d'un nombre infini de perspectives », p 291. Ce problème sera aussi largement abordé par Savage (1954), par exemple p 77. Ici, nous considérerons U comme fini, ce qui aura certaines conséquences, notamment dans la dérivation de la fonction d'utilité.

² À ce sujet voir Granger (1995), p 129-165.

³ « Mixture set », selon la formulation de Herstein/Milnor (1953).

⁴ Voir, Herstein/Milnor (1953), p 292 ; Fishburn (1970) et (1973), p 77 ; Kreps (1988), p 52 et suivantes. La plupart des références postérieures à Herstein/Milnor s'appuient sur leurs démonstrations et leur vocabulaire. N/M n'utilisent pas le concept de « mixture set ».

⁵ Sur ce concept, on peut se reporter à Malinvaud (1952) ; et, Fishburn/Wakker (1995), p 1131-1132.

⁶ La différence entre une classe d'équivalence et une loterie est assez subtile, elle est implicite dans le cadre de l'axiomatique N/M, et elle est importante pour comprendre l'axiome d'indépendance, voir Fishburn/Wakker (1995). Il faut bien comprendre que cet axiome ne peut fonctionner justement qu'entre des classes d'équivalences.

⁷ Plus formellement ; $(\forall u)(\forall v) \in U[(u = v) \vee (u > v) \vee (u < v)]$. Il s'agit d'une *trichotomie*, elle pose des problèmes sur lesquels nous allons revenir. Du point de vue ontologique c'est une opération qui divise le monde en trois possibilités indépendantes, exclusives et non complémentaires.

(3 : A : b) $u > v, v > w$ implique $u > w$ »¹.

La relation $>$ est un ordre total sur U , c'est-à-dire que tous les éléments de U sont ordonnés par $>$. Autrement dit, il n'y a pas d'éléments de U qui ne puissent être comparés². Rappelons que Ramsey lui aussi avait postulé que les sujets soient capables d'ordonner toutes les alternatives possibles³. C'est une clause très forte qui pose de nombreux problèmes, notamment celui de l'incomparabilité de certaines alternatives⁴. L'axiome de complétude implique que toutes les options ou mondes possibles soient comparables⁵. Ce qui rend cette clause acceptable c'est le fait que les options en questions sont de pures entités théoriques, la nature de u, v – ce que ces alternatives pourraient être dans l'épaisseur du réel – comme éléments U n'importe pas⁶.

Par ailleurs, la relation de préférence devra aussi satisfaire les conditions suivantes, qui sont implicites⁷ dans la présentation de N/M :

α) Être asymétrique, c'est-à-dire que si u est préféré à v alors v ne peut pas être préféré à u :

$$(\forall u)(\forall v) \in U [(u > v) \supset \neg(v > u)]$$

C'est une caractéristique fondamentale, elle concerne la cohérence du sujet avec lui-même⁸. C'est précisément sur cette caractéristique que porteront les expériences de Lichtenstein et Slovic dans les années 70⁹. L'asymétrie de la relation $>$ se verra mise en échec

¹ N/M, op. cit., p 26.

² C'est ce qui fait la différence entre un ordre total et un ordre partiel. Dans un ordre partiel, il existe des éléments de l'ensemble de référence qui ne sont ordonnés par la relation $>$, ou toute autre relation possible. Il existe donc des éléments incomparables. Voir Fishburn (1970), p 11-16 et Kreps (1988), p 7-11. Ces deux ouvrages ainsi que Fishburn (1973) contiennent de précieux éléments sur la notion de relation binaire, fondamentale en théorie de la décision.

³ Voir, Ramsey (1926), p170. Cette idée est aussi présente chez Savage (1954), p 16, où le sujet devrait être en mesure de pouvoir ordonner toutes les alternatives possibles concernant l'ensemble de son existence.

⁴ Les incomparables sont des éléments qui ne se situent pas sur le même plan, par exemple une vie humaine et une quantité de monnaie. Il y a aussi l'*incommensurabilité* de certaines alternatives, où les enjeux sont tels que la préférence doit aller de soi, comme dans le pari de Pascal, ou alors qu'une relation de préférence soit impossible à établir.

⁵ Voir Peterson (2009), p 169 qui analyse les objections que l'on peut adresser à cet axiome.

⁶ Saint-Sernin (1973) résume le problème ainsi : parlant de la théorie de N/M : « Ainsi l'interprétation sociologique des préférences humaines en est-elle exclue », p 72. C'est un peu la même chose que ce que nous notions au début : il manque à la théorie une importante dimension philosophique et axiologique.

⁷ Nous basons notre analyse sur les explications que font Fishburn (1970) et Kreps (1988) de cette relation.

⁸ Voir, Saint-Sernin (1973), p 72. Encore une fois il faut noter que cette condition ne fonctionne pas nécessairement selon le contexte temporel dans lequel nous nous plaçons, les préférences, en évoluant peuvent se renverser.

⁹ Voir Lichtenstein et Slovic (1968), (1971) et (1973)

de façon empirique, car Lichtenstein et Slovic vont mettre en place des protocoles expérimentaux prouvant qu'il existe un phénomène de « preference reversal »¹. La violation de l'asymétrie sera aussi une conséquence importante du paradoxe d'Allais.

β) La relation de préférence doit être irréflexive². Aucun élément de U ne peut être strictement préféré à lui même :

$$(\forall u) \in U [\neg(u > u)]$$

γ) La relation de préférence doit être transitive. Ainsi que le notent les auteurs eux-mêmes : « a plausible and generally accepted property »³.

Comme nous l'avons vu dans notre partie sur Ramsey, la propriété de transitivité est importante, notamment en raison de la possibilité de pertes continues dues à des préférences cycliques, ce que nous connaissons sous le nom de « money pump argument »⁴. La transitivité peut être formellement exprimée ainsi :

$$(\forall u)(\forall v)(\forall w) \in U \{[(u > v) \wedge (v > w)] \supset (u > w)\}$$

δ) Enfin, la relation de préférence doit être aussi négativement transitive. Bien que les auteurs ne la mentionnent pas c'est une propriété importante et utile de la relation de préférence car elle permet de garantir la transitivité de l'indifférence. On peut l'énoncer ainsi : pour tout u , v et w de U , si il n'est pas vrai que u soit préféré à v et que v soit préféré à w alors il n'est pas vrai que u soit préféré à w ⁵. formellement :

¹ Voir Slovic/Lichtenstein (1983) et L/S (2006), p 1-40.

² Ce qui pour des raisons évidentes ne peut être le cas de la relation d'indifférence \sim , puisque le contraire serait une violation du principe d'identité. Par contre la relation dite de préférence faible \geq peut être réflexive, comme c'est le cas dans la théorie de Savage par exemple.

³ N/M, op. cit., 27.

⁴ Peterson (2009), p 169, analyse les raisons qui pourraient venir jouer contre la propriété de transitivité. Toutes les relations utilisées en théories de la décision doivent être transitives, à savoir l'indifférence \sim et la préférence faible \geq , nous allons voir sous peu que c'est la transitivité de l'indifférence qui pose le plus de problèmes.

⁵ En fait, pour être tout à fait rigoureux, il faut préciser que ce sont les propriétés de transitivité négative et d'asymétrie qui implique la transitivité simple. On trouve la démonstration complète dans Fishburn (1970), p 12 ; en ce qui concerne la transitivité négative on peut voir Kreps (1988), p 8-9 et Peterson (2009), p 93.

$$(\forall u)(\forall v)(\forall w) \in U \{ [\neg(u > v) \wedge \neg(v > w)] \supset \neg(u > w) \} .$$

Dans le cadre de l'axiomatique de N/M, l'indifférence peut-être définie comme suit : pour tout u et v appartenant à U si u n'est pas préféré à v et que v n'est pas préféré à u alors u est indifférent à v . soit :

$$(\forall u)(\forall v) \in U \{ [\neg(u > v) \wedge \neg(v > u)] \supset (u \sim v) \}$$

La transitivité de la relation d'indifférence stipule que si une option est indifférente à une deuxième option et que cette deuxième option est indifférente à une troisième, alors la première et la troisième option sont indifférentes entre elles. Cette propriété soulève des difficultés que nous allons exposer grâce à un exemple : si un sujet est indifférent entre 1\$ et 1.01\$, alors il sera indifférent entre 1.01\$ et 1.02\$, ainsi qu'entre 1.02\$ et 1.03\$, etc. Il sera donc logiquement indifférent entre 1\$ et 100\$, ou même plus que ça, ce qui n'a pas de sens¹.

Voici ce qu'il en est concernant les caractéristiques de la relation de préférence. Cependant il est très important de noter qu'avec cette relation ainsi caractérisée par la complétude, l'asymétrie et la transitivité négative on peut déjà établir une échelle d'utilité ordinale et un théorème de la représentation de type : $(\forall u)(\forall v) \in U [(u > v) \equiv (V(u) > V(v))]$ ².

b – ordre et combinaison : indépendance et continuité.

+ Indépendance.

« (3 : B : a) $u > v$ implies that $u > \alpha u + (1 - \alpha)v$

(3 : B : b) $u < v$ implies that $u < \alpha u + (1 - \alpha)v$ »³

L'aspect paradoxal de l'axiomatique N/M est que l'on n'y trouve pas un axiome d'indépendance présenté de façon totalement explicite, les auteurs n'utilisent d'ailleurs pas le

¹ Ce type d'exemple est connu sous le nom de « small improvment argument », c'est une objection majeure à la transitivité de l'indifférence. Un exemple de « small improvment » est analysé par Peterson (2009), p 171.

² La preuve de ce biconditionnel est donnée par Peterson (2009), p 95, elle n'est valable que dans le cas où U est fini, sinon il faut faire appel aux preuves développées par Herstein/Milnor (1953).

³ N/M (1947), p 26.

terme « indépendance »¹. Cette absence a été beaucoup discutée². Cependant, si l'on regarde les choses d'un peu plus près on s'aperçoit que les clauses qu'introduisent ici N/M sont bien des clauses d'indépendance au sens « moderne »³. En fait c'est Malinvaud (1952)⁴ qui a montré que ces clauses impliquaient la propriété d'indépendance grâce au concept de classes d'équivalence préalablement introduit. Que signifient-elles ? La clause (3 : B : a) signifie que si une option sûre est préférée à une autre option sûre, alors l'ordre des préférences devra être maintenu dans le cas où l'on passe de la certitude de gagner à la possibilité de gagner⁵. Ce qui est donc en jeu ici c'est le rapport des rétributions avec les probabilités, et ce sera précisément sur ce rapport que porteront les critiques d'Allais (1953) et de Kahneman/Tversky (1979). Avant de passer aux problèmes que soulèvent ces axiomes, interrogeons-nous sur la place qu'ils occupent dans l'axiomatique.

Comme le souligne Fishburn (1970) : « Independence axiom, is regarded by many as the core of expected utility theory, for without it the 'expectation' part of expected utility vanishes »⁶. L'axiome d'indépendance joue donc un rôle majeur : sans lui, on ne peut dériver ni fonction d'utilité cardinale, ni théorème de représentation, car les autres axiomes sont insuffisants. On peut par contre dériver la propriété de l'espérance d'utilité à partir du seul axiome d'indépendance, c'est ce que démontre brillamment Picavet (1996)⁷.

¹ Ils feront allusion à l'axiome d'indépendance de Samuelson dans la préface de l'édition de 1953. Comme nous allons le voir, la propriété d'indépendance est liée au concept de *classe d'indifférence* et à l'opération de mixage entre les éléments de U et les probabilités.

² L'article de référence concernant l'histoire et l'analyse de l'axiome d'indépendance est celui de Fishburn/Wakker (1995). Les auteurs étudient le problème de l'indépendance en théorie de la décision dans trois contextes différents : risque, avec N/M, incertitude avec Savage, et enfin dans le cadre de la théorie du consommateur.

³ Dans le sens que cet axiome prendra en fait immédiatement après la parution de l'édition de N/M (1947), notamment dans les articles de Marschak (1950) et Nash (1950), ainsi que Samuelson (1952) et Herstein/Milnor (1953). Ces points de vue peuvent être résumés ainsi : les éléments communs aux options ne doivent pas être pris en compte dans l'établissement des préférences.

⁴ Il faut noter qu'il manque dans la formulation de N/M une troisième alternative, car trois alternatives sont nécessaires pour avoir une propriété d'indépendance, comme nous le verrons chez Samuelson (1952), et bien entendu chez Savage (1954). Certaines théories affaiblissent l'axiome, ou le suppriment afin d'éviter les critiques. C'est le cas par exemple de Machina (1982).

⁵ Voir Saint-Sernin (1973), p 77 ; et Peterson (2009), p 100.

⁶ Fishburn (1970), p 108 ; Peterson (2009), p 100 ; Fishburn/Wakker (1995), p 1131 et suivantes. Aussi comprenons-nous pourquoi c'est sur cette clause en particulier, dont de PCS de Savage sera la généralisation, que se focaliseront les critiques. Voir, aussi, Friedman/Savage (1952), p 469 ; Machina (1987), p 127-128.

⁷ Comme le note cet auteur, p 160-162, toute la démonstration tourne autour du concept de classe d'équivalence qui avait été mis en avant par Malinvaud (1952) qui démontre que c'est en tant que les loteries peuvent être considérées comme des classes d'équivalences dans un même ensemble U que la propriété d'indépendance de type Samuelson fonctionne. Cependant il faut garder à l'esprit que l'axiome d'indépendance de Samuelson n'est pas encore tout à fait général puisqu'il fait intervenir la relation d'indifférence avec les probabilités suivantes : $(\forall u)(\forall v)(\forall w) \in U[(u \sim v) \supset [(u1/2) + (1/2w) \sim (v1/2) + (1/2w)]]$. C'est une version plus faible que la formulation sur laquelle travaille Picavet. Voir aussi Fishburn/Wakker, p 1132 – 1133.

Revenons au texte et aux explications que proposent les auteurs à propos des deux clauses d'indépendance : « This is legitimate since any kind of complementarity (or the opposite) has been excluded »¹. Ces clauses sont donc reliées à la propriété suivante : les événements sont exclusifs et ne peuvent être réalisés en même temps², ni « déborder » l'un sur l'autre, il s'agit en fait de respecter les axiomes du calcul des probabilités et notamment celui-ci : $P(p \vee q) = P(p) + P(q)$ ³.

Les événements, pondérés par α et $(1 - \alpha)$, ne sont ni complémentaires, ni substituables, ils sont strictement indépendants. Pourtant, si nous regardons les choses plus en profondeur une autre explication est plausible, explication qui fait à nouveau appel au concept de « classe d'équivalence » définit plus haut dans l'axiomatique, soit la formulation canonique suivante avec trois alternatives :

$$(\forall u)(\forall v)(\forall w) \in U (\forall \alpha) \in [0,1] [(u > v) \supset (u\alpha + (1 - \alpha)w > v\alpha + (1 - \alpha)w)]$$

Ici nous pouvons analyser l'axiome comme signifiant que u est une classe d'équivalence de la loterie $u\alpha + (1 - \alpha)w$ dans U , et que v est une classe d'équivalence de la loterie $v\alpha + (1 - \alpha)w$ dans U , ce qui explique de façon immédiate l'équivalence entre les options sûres et les loteries d'une part, et par là même le fait que l'ordre des préférences soit préservé lorsque l'on passe du membre de gauche au membre de droite. Une des propriétés majeures de cet axiome est qu'il indique que, dans une situation décisionnelle donnée, le sujet doit se focaliser sur les éléments qui rendent les alternatives différentes, et ne pas prendre en compte les éléments qui sont semblables⁴ ici : $(1 - \alpha)w$.

Pour prendre un exemple simple disons que si un sujet préfère une poire à une pomme, il devra préférer la loterie « avoir une poire avec la probabilité α , une orange avec la probabilité $1 - \alpha$ » à celle « avoir une pomme avec la probabilité α , une orange avec la probabilité $1 - \alpha$ ». La partie commune aux alternatives, qui se réalise systématiquement si α n'arrive pas, ne doit pas être prise en compte et ne doit pas influencer l'ordonnement des

¹ N/M, op. cit., p 27.

² N/M, op. cit., p 18 ; et Fishburn/Wakker (1995), p 1132.

³ Voir Kolmogorov (1933) ; Fishburn/Wakker, ibid.

⁴ Le Principe de la Chose Sûre (PCS) de Savage généralisera l'axiome d'indépendance au contexte d'ignorance en montrant que les préférences doivent être indépendantes de la réalisation de certains événements ; voir Savage (1954), p 26.

préférences¹.

L'axiome d'indépendance soulève la délicate question du rapport entre préférences et probabilités². En effet, la préférence originelle doit être maintenue quelle que soit la valeur de α , c'est-à-dire, en reprenant notre exemple, je dois préférer la loterie qui me donne 0,01 chance d'avoir une poire et 0,99 chances d'avoir une orange à celle qui me donne 0,01 chances d'avoir une pomme et 0,99 chances d'avoir une orange. Cela s'explique par le fait que justement nous avons dans les deux loteries les mêmes distributions de probabilités, et c'est pour cette raison que l'axiome d'indépendance peut s'appliquer, car si les probabilités étaient différentes et si nous avions par exemple 0,01 chance d'avoir une poire et 0,99 chance d'avoir une orange contre 0,99 chances d'avoir une pomme et 0,01 chance d'avoir une orange, alors il serait peut-être *plus rationnel* de préférer avoir une pomme. C'est ce que va mettre en avant le « paradoxe » d'Allais : les sujets établissent leurs préférences en partie en fonction des distributions de probabilités, et lorsque les probabilités s'approchent de la certitude, les gains qui leur sont attachés passent en première position dans l'ordre des préférences.

+ Continuité.

« (3 : B : c) $u < w < v$ implies the existence of an α with

$$\alpha u + (1 - \alpha)v < w$$

(3 : B : d) $u > w > v$ implies the existence of an α with

$$\alpha u + (1 - \alpha)v > w \text{ »}^3$$

Il s'agit de deux clauses de continuité, comme le note Gilboa (2009), elles jouent un rôle important dans la structure mathématique de la théorie, cependant elles sont insuffisantes à elles seules pour dériver une fonction d'utilité⁴. Rappelons que Ramsey avait lui aussi introduit ce type d'axiome sous la dénomination d'« axiome d'Archimède », il avait pour but de garantir que toutes les différences de valeurs entre les options ou mondes possibles puissent

¹ Les exemples de ce type abondent dans la littérature, mais celui de Peterson (2009) basé sur le paradoxe d'Allais est particulièrement pertinent, puisqu'il permet de mettre en évidence l'aspect normatif du principe d'indépendance. C'est autour de la normativité de ce principe que vont se structurer la plupart des critiques que nous étudierons dans la prochaine partie.

² C'est précisément l'inverse que semblent soutenir Friedman/Savage (1952), p 467.

³ N/M, op. cit., p 26.

⁴ Voir, Gilboa (2009), p 81.

être commensurables¹. Nous allons observer que les clauses proposées par N/M sont assez proches de l'intuition de Ramsey, même si pour ce dernier le ressort de l'axiome se situait dans la notion d'égalité alors que pour N/M il s'agit de montrer que certaines distributions de probabilités² peuvent modifier l'ordre des préférences entre trois options et rendre aussi peu désirable que possible une options qui dans le choix original était située en première position. Quel est le sens de ces deux clauses, que nous allons résumer dans la formule suivante :

$$(\forall u)(\forall v)(\forall w) \in U (\exists \alpha)(\exists \beta) \in [0,1] \{ (u > w > v) \supset [u\alpha + (1-\alpha)v] > w > [u\beta + (1-\beta)v] \}$$

Cet axiome formalise le fait que les préférences peuvent être soumises à des variations se situant en continuité avec les probabilités, ce qui est une idée centrale et maîtresse de la TUA³. Comme l'indiquent les auteurs : « However desirable v may be in itself, one can make its influence as weak as desired by giving it a sufficiently small chance »⁴.

Voyons ce que cette définition implique. Elle formalise le fait qu'une option sûre, aussi bonne soit-elle (u dans notre formule) en elle-même dans l'ordre des préférences, peut voir sa désirabilité diminuer en fonction de la distribution de probabilité qui lui est appliquée lorsqu'elle est présentée sous la forme d'une loterie dont le deuxième membre est la pire alternative (v dans notre formule). Symétriquement l'option sûre (w) qui vient en dernier dans les préférences du sujet, peut voir sa désirabilité augmenter par rapport à la préférence médiane (v) si elle devient le membre d'une loterie, dont l'autre membre est la meilleure alternative (u), et qu'elle est pondérée par une probabilité très faible.

Prenons un exemple. Soit $u = 100\$$, $w = 50\$$, et $v = 0\$$. L'axiome de continuité nous apprend qu'il existe deux distributions de probabilités⁵ α et β comprises dans l'intervalle $[0,1]$ qui rendent la loterie $[100\$\alpha + 0\$(1-\alpha)]$ préférable – avec α proche de 1 – à l'option certaine 50\$. Mais, inversement, une distribution de probabilités β appliquée à la loterie

¹ Voir Ramsey (1926), p 171.

² Il est intéressant de remarquer que chez N/M, c'est la distribution de probabilité qui joue le rôle dévolu au nombre entier n dans l'axiome d'Archimède. C'est donc la probabilité qui va servir d'unité de mesure entre les différentes options.

³ Voir, Friedman/Savage (1952), p 467 ; Binmore (2009), p 39 ; et, Mongin (2011), p 3. Cette propriété est aussi comprise dans la clause 2) que nous avons présenté plus haut. Comme nous le notions, il semble y avoir une tension entre l'axiome d'indépendance et l'axiome de continuité. Nous reviendrons sur cette tension en conclusion.

⁴ N/M, op. cit., p 27.

⁵ C'est là, une différence fondamentale d'avec l'axiome d'indépendance, dans lequel les distributions de probabilités sont les mêmes dans les deux loteries.

$[100\beta + 0(1 - \beta)]$ peut rendre l'option certaine 50\$ préférable quand β est proche de 0¹. L'aspect normatif de l'axiome de continuité a été lui aussi beaucoup discuté², surtout lorsque nous mettons en jeu des options ou des perspectives extrêmes dans l'échelle des biens et des maux, comme la mort par exemple. Pourtant il semblerait que c'est l'axiome de la TUA qui a l'aspect normatif le moins discutable. Ici encore, nous allons user d'un exemple.

Considérons les options suivantes : $u = 1000\$$, $w = 100\$$, et $v = \text{être exécuté immédiatement}$. Il est bien naturel, avec ces alternatives d'avoir l'ordre de préférences suivant : $u > w > v$ et de se dire que quelle que soit la distribution de probabilités sur la loterie $[u\alpha + (1 - \alpha)v]$ jamais je ne prendrai la plus petite chance de jouer ma vie contre 1000\$ ou n'importe quelle somme d'argent, aussi grande soit-elle d'ailleurs. Modifions un peu les données du problème de façon à faire l'expérience suivante : on propose au sujet de gagner 100\$ tout de suite ou de prendre sa voiture, rouler 5 kilomètres, et récupérer une enveloppe contenant 1000\$. Le choix du décideur serait ici tout à fait différent, et l'axiome fonctionne parfaitement, aussi bien d'un point de vue normatif que descriptif. La réticence originale vient de *l'incommensurabilité théorique* entre les deux alternatives. Mais après réflexion il devient clair que dans la réalité, les chances de mourir, en effectuant une action susceptible de maximiser un bien ou des avoirs, ne sont jamais nulles.

Nous pourrions même aller plus loin en disant que nous vérifions quotidiennement la pertinence de cet axiome, par exemple dès que nous traversons la rue pour acheter un journal ou autre, bref dès que nous agissons. L'axiome de continuité se voit ainsi doté d'une véritable portée philosophique et d'une composante existentielle non négligeables : tout simplement le fait que dans nos vies, nos chances de mourir ne sont jamais nulles. Ce qui nous amènera dans notre troisième partie à tenter de répondre à cette question : notre finitude doit-elle être

¹ Voir Peterson (2009), p 100. Il est important de remarquer que cet axiome contient, de façon virtuelle, l'intuition de l'effet certitude (« certainty effect ») qui sera mis en avant par Allais (1953) et Kahneman/Tversky dans la Prospect Theory (1979). En effet, la certitude joue ici un rôle central dans la comparaison de l'option certaine avec $[u\beta + (1 - \beta)v]$, l'idée est la suivante : un gain moyen est systématiquement préféré à une probabilité assez haute d'un gain nul, cependant nous allons voir que d'un point de vue descriptif et normatif les choses sont plus nuancées. Par contre, l'effet certitude, ne joue pas de rôle dans la comparaison entre $[u\alpha + (1 - \alpha)v]$ et l'option sûre, car nous postulons que la probabilité d'avoir l'option la plus intéressante est proche de 1. Il serait intéressant d'un point de vue expérimental de déterminer à partir de quelles distributions de probabilités le sujet devient indifférent entre les loteries et l'option médiane sûre.

² À ce sujet, on peut se reporter à Chernoff (1959), p 81-82 ; Fishburn (1973), p 109-110 ; Kreps (1988), p 44-46 ; Peterson (2009), p 100 et suivantes. Cependant comme cet axiome est moins central en ce qui concerne l'utilité attendue, il a été un peu moins systématiquement étudié. C'est un axiome qui peut d'ailleurs être remplacé, c'est ce qu'a fait Hausner (1954) qui utilise à la place de cet axiome une clause d'indépendance affaiblie, basée sur l'indifférence, probablement inspirée de Samuelson (1952).

intégrée – sur le mode de la résistance – de façon systématique à l'établissement de nos buts et visées, ainsi qu'à notre système de référence décisionnel?

c – Combinaison et algèbre.

$$\ll (3 : C : a) \quad \alpha u + (1 - \alpha)v = (1 - \alpha)v + \alpha u$$

$$(3 : C : b) \quad \alpha(\beta u + (1 - \beta)v) + (1 - \alpha)v = \gamma u + (1 - \gamma)v$$

$$\text{Where } \gamma = \alpha\beta \gg^1$$

La première des deux clauses va de soi puisque nous avons à faire à des événements alternatifs. La deuxième clause – réduction des loteries composées à une loterie simple – est plus importante, elle doit faire l'objet de quelques remarques². Tout d'abord cette clause garantit l'applicabilité du calcul des probabilités à des loteries d'une complexité quelconque, aussi grande que l'on veut. Deuxièmement, cet axiome signifie que toute loterie composée peut être réduite à une loterie simple selon les règles du calcul des probabilités. Si d'un point de vue théorique l'opération est parfaitement légitime, d'un point de vue pratique il n'en est pas ainsi, comme le remarquent d'ailleurs les auteurs, car cet axiome : « gets closest to excluding a 'utility of gambling' »³. Cette clause échoue à formaliser un phénomène important qui peut être qualifié de « plaisir du jeu », en postulant, comme le note Picavet (1996) un « ordre complet pour les tickets composés, qui coïncide avec l'ordre des tickets simples auxquels on peut les réduire »⁴. Du point de vue du comportement il n'est pas sûr en effet que le plaisir pris à un double ou un triple tirage soit le même que celui pris à un tirage unique⁵.

C'est ce que mettront aussi en avant les critiques d'Allais (1953). La loterie simple aura bien entendu la même utilité attendue que toutes les loteries prises séparément et diachroniquement. Il reviendra à Kahneman et Tversky, dans leur article séminal de 1979 sur

¹ N/M, op. cit., p 26.

² Comme le note Peterson (2009), il s'agit plutôt d'une clause technique qui permet de simplifier la présentation des options. Fishburn (1973) et Kreps (1988) ne mentionnent pas cette propriété, car elle ne joue aucun rôle dans la dérivation de la fonction d'utilité. Luce/Raiffa (1957) donnent une explication claire de cet axiome, p 26-27, et proposent des remarques sur l'aspect normatif versus l'aspect descriptif de cette clause.

³ N/M, op. cit., p28. Cet aspect de l'axiome de réduction des loteries composées fera l'objet de différentes remarques dans Samuelson (1952), p 671, et dans Allais (1952), (1953).

⁴ Picavet (1996), p 159.

⁵ Le roman *Le joueur* de Dostoïevski (1865) nous offre une description particulièrement pertinente du phénomène de « fièvre du jeu ».

la Prospect Theory, de remettre définitivement cette clause en question, en montrant que les sujets n'évaluent pas de la même façon les tirages simples et les tirages successifs, même s'ils ont la même utilité attendue¹.

Ainsi, à partir des axiomes introduits : « We can derive (...) the numerical character of utility »². Et ce, dans le sens qui avait été établi plus haut, c'est-à-dire que les utilités doivent pouvoir être mises en correspondance avec des nombres réels d'une part : $u \rightarrow p = V(u)$. C'est le problème de la mesure de l'utilité, dont cette correspondance est la condition de possibilité. D'autre part, en respectant la clause 1) l'utilité attendue doit respecter l'ordre : $(u > v) \equiv (V(u) > V(v))$ et il est intéressant de noter que c'est justement cette correspondance entre u et p qui garantit l'applicabilité de la relation $>$ aux utilités. Enfin, V , doit être une fonction linéaire par rapport aux probabilités, c'est-à-dire avoir la propriété d'expectation en respectant la clause 2) : $V(\alpha u + (1-\alpha)v) = \alpha V(u) + (1-\alpha)V(v)$. Ainsi : « The numerical values of utility combine (with probability) like mathematical expectation ! »³. En fin de compte c'est la conjonction des trois axiomes : ordonnancement, indépendance et continuité qui permet de dériver la fonction d'utilité⁴. Pour être plus exact il faut dire que la relation $>$ satisfait l'ordonnancement, la transitivité, l'indépendance et la continuité *si et seulement si*⁵ il existe une fonction u de U dans \mathbb{R} ⁶ qui satisfait les propriétés que nous avons mentionnées :

$$1) (u > v) \equiv (V(u) > V(v))$$

$$2) V(\alpha u + (1-\alpha)v) = \alpha V(u) + (1-\alpha)V(v)$$

Ce sont les deux propriétés que nous avons étudiées en premier lieu, comme prérequis à l'axiomatique : elles définissent les caractéristiques de la fonction d'utilité. Cependant pour

¹ Voir, Kahneman/Tversky (1979), p 25-27.

² N/M, op. cit., p 28.

³ N/M, ibid.

⁴ Pour plus de détails, voir, Fishburn (1970), p 107-108 ; Kreps (1988), p 46.

⁵ Cette nuance est très importante, car elle signifie que nous pouvons aller dans les deux sens, c'est-à-dire des axiomes vers la fonction d'utilité, et réciproquement de la fonction d'utilité aux axiomes. C'est d'ailleurs en suivant cette double direction que les auteurs présentent leurs dérivations. Voir, entre autres, Chernoff (1959), p 350-352 ; Peterson (2009), p 301-307. La preuve de Peterson reste la plus pédagogique et la plus claire.

⁶ Fonction dont les arguments sont des loteries.

que u soit unique, une troisième propriété est requise, à savoir que toutes les fonctions satisfaisant 1) et 2) soient des transformations linéaires positives les unes des autres. C'est-à-dire que pour toute fonction u' , autre que u , satisfaisant 1) et 2), il existe deux nombres, $c > 0$ et d , tels que : $u' = cu + d$. Cette propriété, combinée à celles qui précèdent et aux axiomes permet à N/M d'établir que : « We have practically defined numerical utility as being that thing for which the calculus of mathematical expectation is legitimate »¹

Ce sont donc bien les idées d'expectation et de maximisation, mathématiquement construites et définies, qui constituent l'essence même de la TUA, et ce en continuité avec les intuitions de Ramsey, même si l'interprétation des probabilités change avec N/M ; il reviendra à Savage d'opérer la synthèse originale et définitive entre les points de vue de N/M et celui de Ramsey et de Finetti. D'autre part, c'est à propos des idées d'expectation et de maximisation que vont bientôt se centrer les critiques qui viendront de l'économie avec Allais, et de la psychologie avec Edwards.

En conclusion nous voudrions aborder brièvement un dernier point important : que signifie la tension qui semble être à l'œuvre entre l'axiome d'indépendance et l'axiome de continuité. D'autre part, nous nous interrogerons sur le rapport de la théorie de l'utilité avec la « psychologie populaire », et enfin sur le problème de la normativité de la TUA.

Comme nous le remarquons, à première vue, l'axiome d'indépendance et l'axiome de continuité semblent donner lieu à des interprétations opposées. Le premier affirme que l'ordonnancement des préférences ne doit être altéré ni par les éléments communs aux options, ni par un changement dans la distribution des probabilités, par exemple le passage de la certitude à la possibilité. D'autre part l'axiome de continuité nous apprend que deux distributions de probabilités différentes peuvent modifier la valeur absolue des options, et rendre par exemple la pire option préférable à l'option médiane si elle est incorporée dans une loterie et que ses probabilités sont faibles. À première vue il semblerait que ces deux interprétations soient en contradiction l'une avec l'autre. Or il n'en est rien. Les propriétés de ces axiomes sont liées aux distributions de probabilités que nous utilisons. Dans le cas de la

¹ N/M, op. cit., p 28.

clause d'indépendance, la distribution de probabilité est la même sur les deux options, car dans l'hypothèse contraire, nous ne pourrions plus parler d'éléments identiques dans les alternatives : les événements seraient pondérés par des probabilités différentes, les alternatives seraient donc différentes et l'axiome ne pourrait pas fonctionner. Dans le cas de la continuité nous avons le phénomène inverse : si les loteries n'étaient pas pondérées par des probabilités différentes alors l'axiome ne fonctionnerait pas et n'aurait aucun sens, car nous aurions alors une loterie qui serait préférée à une alternative sûre qui serait à son tour préférée à la loterie initiale.

Enfin, une question cruciale se pose concernant l'axiomatique N/M, question qui peut être généralisée à toutes les tentatives de formalisation de l'utilité et de la décision : quel est le rapport de cette théorie avec la psychologie populaire ou ordinaire (« folk psychology »¹) ? Dans quelle mesure la TUA s'inscrit-elle en continuité avec la psychologie populaire dont elle serait, en quelque sorte, la mise en forme ? Dans quelle mesure la TUA peut-elle « améliorer » la psychologie ordinaire, c'est-à-dire la manière dont nous décidons et jugeons habituellement ? Ce sont des questions que se sont par exemple posés des auteurs comme Davidson et Lewis².

En fin de compte, la question est la suivante : dans quelle mesure la TUA peut-elle être considérée comme normative ou descriptive ? C'est une question importante, car elle met en cause l'applicabilité de la théorie elle-même. Il faut donc se poser la question suivante : la TUA peut-elle être appliquée à tous les types de décisions ? Nous verrons que Savage répondra à cette question par la négative. C'est une théorie très délimitée dans le temps et dans l'espace, et c'est peut-être ici d'ailleurs que nous entrevoyons une différence fondamentale entre N/M et Savage : chez les premiers le domaine d'applicabilité de la théorie semble plus large, puisque la théorie est censée pouvoir s'appliquer au contexte économique ; la théorie de Savage sera elle bien plus limitée dans son applicabilité, d'ailleurs, à aucun moment Savage n'envisage la possibilité que sa théorie puisse être appliquée à un système aussi vaste que l'économie.

¹ Sur ce concept on se réfère à Stich (1983).

² Voir, Davidson (2001), chapitres 12 et 14, et Lewis (1983) et (1986) qui pense que la TUA est une systématisation de la psychologie ordinaire. Pour Davidson la TUA permettrait de mieux expliquer le choix rationnel, en connexion avec les concepts de désir et de croyance. C'est la fonction explicative de la TUA qui est centrale pour cet auteur.

C- Décision et incertitude : la théorie de Savage.

1) Les définitions préliminaires et le découpage ensembliste.

La théorie de N/M bien qu'étant très achevée, nous laisse avec un problème de taille : comment analyser une situation décisionnelle lorsque les probabilités ne sont pas objectives¹ ? Comment formaliser la prise de décision en contexte d'incertitude, c'est-à-dire lorsque les probabilités ne sont pas données objectivement, mais construites par le sujet à partir de ses croyances² ? C'est la question à laquelle, dans la lignée de Ramsey (1931), va répondre de façon magistrale le mathématicien L. J. Savage en rédigeant son maître ouvrage : *The foundations of statistics*³. L'axiomatique de Savage représente une avancée majeure, probablement même l'aboutissement majeur en théorie statistique de la décision, et le dépassement des travaux fondateurs de Neyman, Pearson et Fisher. Nous pouvons même dire que c'est la théorie bayésienne de la décision qui reste à ce jour la plus achevée⁴. Avant Savage, la théorie bayésienne du choix rationnel était presque totalement ignorée⁵, après lui,

¹ Voir, Gilboa (2009), p 40-46 ; et l'article de R. Sugden, in Barberà/Hammond/Seidl (2006), p 726. Sur la dualité « probabilités subjectives et probabilités objectives » en théorie de la décision on peut se reporter à Saint-Sernin (1973), p 38-60. L'auteur recense les efforts de ceux qui depuis de Finetti (1937) et Carnap (1945) ont tenté d'unifier les deux points de vue, que ce soit sous l'égide du subjectivisme ou d'une approche plus logiciste. Savage (1954) ira, comme nous le verrons, dans le sens de de Finetti. Il faut toutefois noter que d'autres auteurs ont tenté d'unifier les deux points de vue en les incorporant dans une seule théorie de l'utilité, c'est le cas par exemple de Anscombe/Aumann (1963), et de Raiffa (1968).

² Comme le fait remarquer Picavet (1996), p 207, une des originalités de la théorie de Savage, c'est le rôle opératoire qu'elle attribue au concept de croyance.

³ L'ouvrage a été publié en 1954, et a connu une deuxième édition en 1972, avec certaines corrections et ajouts bibliographiques dont certains sont majeurs en théorie de la décision par exemple : Edwards/Savage/Lindman (1963), Ellsberg (1961) ou Hacking (1967). Selon les mots de l'auteur, p 5, il se divise en deux parties : chapitre 2 à 7 pour l'aspect fondationnel proprement dit, et chapitres 8 à 17 pour l'élaboration complète de la théorie avec la reprise de la théorie du minimax au chapitre 9, et l'hypothèse du « minimax regret ». C'est un livre très difficile qui a pourtant donné lieu à une abondante littérature secondaire dans le domaine des mathématiques de la décision, de l'économie et du bayésianisme. Cependant il faut prendre garde que la complexité de l'ouvrage a donné lieu à deux types d'approches : scientifiques et vulgarisatrices. Dans le premier cas nous avons par exemple des ouvrages de références comme Fishburn (1970), Kreps (1988), Barberà/Hammond/Seidl (2006). Concernant le point de vue économique, ou plutôt microéconomique, car il est impossible d'appliquer la théorie de Savage à un ensemble aussi vaste que l'économie à grande échelle, on peut se reporter à Fienberg/Zellner (1975). Du côté français, la théorie de Savage a été étudiée par Saint-Sernin (1973), qui consacre la majeure partie de son analyse au problème de la normativité des probabilités ; Granger (1988) consacre une brève mais très pertinente partie à Savage dans le cadre d'une enquête sur le « style » des interprétations subjectivistes en probabilités ; Picavet (1996) propose l'analyse la plus exhaustive et la plus précise disponible en langue française. En ce qui concerne les ouvrages de vulgarisation (presque tous excellents) nous avons le classique de Luce/Raiffa (1957) ; Baron (1988) ; Peterson (2009) ; Binmore (2009) ; et enfin Gilboa (2009), le livre de Gilboa se situe en fait à mi-chemin entre l'approche scientifique et l'approche plus « populaire ». Le livre de Savage a donné lieu à bon nombre d'articles que nous mentionnerons à l'occasion.

⁴ Kreps (1988) va jusqu'à qualifier l'entreprise de Savage de « The crowning glory of choice theory », p 120. Voir aussi Picavet (1996), p 190-191.

⁵ Voir à ce propos le court mais intéressant article de Savage : « The shifting foundations of statistics » dans R. G. Glodny (Ed.) (1977), p 3-17.

elle occupera une place de premier plan dans les débats scientifiques concernant les probabilités personnelles¹ et certaines parties de l'épistémologie contemporaine².

Nous allons ici étudier les cinq premiers³ chapitres de l'ouvrage et suivre d'aussi près que possible le déploiement de la théorie avec l'introduction des 7 axiomes⁴ qui vont permettre d'établir *du même coup* l'existence d'une probabilité qualitative sur les événements et l'existence d'une fonction d'utilité obéissant au principe bernoullien de l'espérance mathématique⁵, fonction d'utilité qui sera construite à l'aide du concept de « pari »⁶. Le livre de Savage, en ce sens qu'il propose une théorie de l'utilité fondée sur des probabilités subjectives, peut être considéré comme la synthèse des travaux de Ramsey (1926), de Finetti (1937)⁷ et von Neumann/Morgenstern (1947).

Après quelques remarques d'ordre philosophiques sur les *fondations* dans les sciences dures⁸, Savage présente immédiatement l'interprétation des probabilités⁹ qu'il entend soutenir et développer au long de son ouvrage :

« Probability measures the confidence that a particular individual has in the truth of a particular proposition, (...). These views postulate that the individual concerned is in some way 'reasonable' »¹⁰.

Deux remarques s'imposent. Tout d'abord l'auteur relie de façon immédiate la rationalité de l'agent à l'application et au respect des axiomes du calcul des probabilités¹¹,

¹ Voir, Savage (1961), (1967) et (1968) ; ainsi que Kyburg/Smokler (1980).

² Voir, Bovens/Hartmann (2003), et Oaksford/Chater (2007).

³ Ils se situent aux pages 1-104, nous citons : Savage (1954).

⁴ Savage utilise le terme « postulate ». Certains théorèmes sont très importants aussi.

⁵ Voir, Savage (1954), p 74-76.

⁶ La filiation entre Savage et Ramsey est très nette, et parfaitement assumée par notre auteur, voir, Savage (1954), p 7 ; et Picavet (1996), p 191.

⁷ Savage se réfère beaucoup aux résultats de l'article de 1937, surtout dans les chapitres 3 et 4 sur les probabilités personnelles, en particulier p 40 à 56 où Savage explore à sa façon le concept « d'échangeabilité ».

⁸ Du point de vue épistémologique, Savage semble être assez en accord avec les idées de Quine (1951) ; par ailleurs l'inspiration générale de l'ouvrage, et notamment le rôle attribué à la croyance, pourrait laisser supposer une ascendance humienne assez forte, nous verrons, que si l'inspiration est certes humienne, la théorie de Savage ne peut en aucun cas être considérée comme la mise en forme mathématique de l'ensemble des idées de Hume sur les croyances. Voir, Picavet (1996), p 207.

⁹ L'auteur discute aussi brièvement, trop brièvement peut-être, les autres interprétations. Pour un point de vue plus complet de Savage lui-même, on peut se référer à l'article cité en note 5.

¹⁰ Savage (1954), p 3.

¹¹ Il faut insister brièvement sur l'aspect normatif des probabilités, conçues par Savage comme « un code de cohérence » pour le sujet. Nous sommes ici très proches des intuitions de Ramsey (1926), qui faisait des probabilités les « principes de la raison humaine », d'autre part, toujours dans la lignée du philosophe britannique il s'agit d'une croyance en la vérité d'une

probabilités conçues comme une mesure des degrés de croyance. Il s'agit ici, comme chez Ramsey (1926) et de Finetti (1937) de construire une structure théorique qui puisse relier les probabilités au *comportement* observable des sujets en contexte décisionnel.

Deuxième remarque : nous assistons à un écart avec la théorie de N/M. En effet deux agents « rationnels » dans la même situation, avec les mêmes informations peuvent développer des degrés de croyance différents et des préférences différentes. Le postulat implicite de Savage est le suivant : nos croyances arrivent par degrés et elles ont une structure probabiliste qui permet d'en formaliser les modifications¹. La théorie de Savage est, comme celle de Ramsey, à connecter avec la théorie humienne de la croyance comme disposition à l'action², c'est ce que viendra formaliser, entre autres l'axiome 4³.

Pour Savage, les décisions en contexte d'incertitude (« in the face of uncertainty ») sont le lot quotidien des sujets humains, des organisations et même des animaux⁴. En effet, il y a un nombre extrêmement élevé de variables que nous ne contrôlons pas dans notre environnement décisionnel. La question qui se pose ici est de savoir dans quelle mesure la logique – ici une logique inductive du type de celle que propose Ramsey⁵ – peut pallier cet inconvénient : « It must be admitted that logic itself does lead some criteria for decisions »⁶.

Si la logique de la décision peut être conçue comme un guide pour le choix rationnel et par là même pour l'action, nous pouvons nous demander quel type de « décideur » Savage a à l'esprit lorsqu'il élabore sa théorie : « I am about to build up a highly idealized theory of the behavior of a 'rational' person with respect to decision »⁷. L'agent rationnel auquel s'adresse

proposition. Une de nos analyses dans notre troisième chapitre insistera justement sur la pertinence de ce « code de cohérence » dans le cadre du fonctionnement du système de référence décisionnel. Concernant l'aspect normatif des probabilités, voir Savage, op. cit., p 59 ; et aussi l'excellente analyse de Saint-Sernin (1973), p 45 ; et Picavet (1996), p 191. Pour un point de vue plus récent sur cette problématique, voir T. Martin (Ed.) (2003).

¹ C'est un des réquisits centraux du bayésianisme ; voir, Gärdenfors (1988), p 4-10 ; Baron (1988), p 121-134 ; Gilboa et Al. (2004), p 3-4 ; Griffith/Yulle in Chater/Oaksford (2008), p 33-57. Peterson (2009), p 125-130 ; Binmore (2009), p 116-134. On peut aussi se reporter à Bradley (2006), p 5-6 qui récapitule et unifie les vues de Ramsey et Savage.

² Voir à ce sujet Engel (1997) et Picavet (1996), p 53-95.

³ L'étude de cet axiome nous révélera que les choses sont plus nuancées et plus complexes.

⁴ Savage associe plusieurs fois les animaux aux problèmes décisionnels.

⁵ Voir Ramsey (1926), p 187.

⁶ Savage (1954), p 6. Il est intéressant de comparer cette phrase avec ce que dit Savage sur le rôle « régulateur » de la logique à la page 59 : qu'elle ne saurait être le seul critère normatif pour l'existence. Il y a une sous-détermination radicale de la vie par la logique, et ce dans de multiples dimensions qui dépassent largement le cadre de la théorie de la décision. C'est ce que nous tenterons de montrer dans notre troisième chapitre : la philosophie de la volonté doit remplacer la logique de la décision. Concernant l'aspect de la logique conçue comme un ensemble de règles à caractère normatif et les problèmes qui y sont reliés, on peut se reporter aux travaux séminaux de Wason (1960), (1966) ; ainsi qu'à Baron (1988), p 50 et p 77-97. Une partie de l'entreprise de Kahneman et Tversky, dans leur théorie des heuristiques et des biais, sera de montrer que la logique inductive, non plus n'est pas normative ; voir, Tversky/Kahneman (1974) ; Andler (2004), p 315-405.

⁷ Savage (1954), p 7.

Savage est une *personne hautement idéalisée*, c'est-à-dire capable de faire des choix totalement logiques et cohérents¹. En outre, l'agent en question, comme c'était déjà le cas chez Ramsey et N/M, doit être capable d'ordonner et de hiérarchiser toutes les options qui se présentent à lui. La personne idéale à laquelle se réfère Savage est celle qui respecterait ses axiomes dans toutes les situations possibles, pourvu qu'elles soient assez circonscrites et précises².

La première étape de cette reconstruction du choix rationnel consiste en un découpage aussi fin que possible de la réalité décisionnelle en termes d'ensembles, de sous-ensembles, d'éléments et d'algèbres de sous-ensembles³. Savage introduit deux ensembles fondamentaux – que viendront relier des fonctions : les actes – avec leurs éléments :

- *S* : le monde. Mais il faut être attentif au fait que le monde en question est l'ensemble de tous les éléments *pertinents* pour une décision donnée, c'est-à-dire les états du monde possibles dans un contexte précis. Il ne s'agit donc pas ici de l'ensemble de tous les objets de l'univers⁴. La taille du monde à considérer est tributaire du type de décision que nous avons à prendre. Donc : « In application of the theory, the question will arise as to which world to use in a given context »⁵.

Par exemple, si le sujet a en face de lui une douzaine d'œufs et qu'il se demande si l'un d'eux ou plusieurs sont avariés, le monde pertinent dans ce contexte-là sera l'ensemble de tous les sous-ensembles de descriptions possibles attachées à la douzaine d'œufs, soit 2^{12}

¹ Remarquons que Savage soulève ici un problème important, qui va être dirimant en théorie de la décision : dans quelle mesure la théorie peut-elle avoir une portée pratique, ou plutôt, dans quelle mesure la théorie est-elle applicable en contexte décisionnel par les sujets que nous sommes ? À ce sujet, et pour une approche originale de cette question, voir Weirich (2004). L'ouvrage de cet auteur, dont le sous-titre est : « Rules for non-ideal agents in non-ideal circumstances » semble être une réponse partielle au problème soulevé ici par Savage. Une large partie des travaux séminaux de Gigerenzer, sur lesquels nous reviendrons, seront consacrés à la critique des postulats de rationalité dans le cadre d'une analyse de la rationalité limitée.

² Pour Savage il y a une adéquation stricte entre l'axiomatique et la rationalité, nous pourrions même dire que la rationalité de l'agent passe par le respect de l'axiomatique. Toutefois, nous pouvons d'ores et déjà noter que le « code de cohérence » pour un sujet idéal que constitue la théorie de Savage sera la cible de nombreuses critiques, dont celle d'Allais (1953). Pour un point de vue plus contemporain sur la normativité des axiomes de Savage, voir Gilboa et Al. (2009) ; ainsi que Picavet (1996), p 222-228 qui s'appuie sur les critiques importantes formulées contre la théorie de Savage par R. Sugden (1991).

³ Pour une introduction à ces notions, voir Fishburn (1973), p 130-134 ; Kreps (1988), p 115-126.

⁴ Savage envisage cette possibilité page 9, mais il la rejette immédiatement, car il la trouve trop vague et peu réaliste : les mondes dans lesquels nos décisions prennent place doivent être le plus circonscrits possible, et se situer dans ce que Savage nomme un « petit monde », l'auteur analyse ce concept aux pages 82-90 ; voir aussi Picavet (1996), p 196 et suivantes ; Shafer (1986), p 466-467.

⁵ Savage (1954), p 9. Comme nous le remarquons dans la note précédente, cela nous indique que la théorie de Savage est d'*emblée* limitative, elle ne pourra être appliquée qu'à des décisions inscrites dans un « petit monde ».

descriptions possibles. Le concept de monde introduit par Savage est un ensemble d'états du monde qui sont ses éléments et qui sont notés s . Parmi ces états du monde possibles, un seul est vrai, un seul va se réaliser à l'exclusion¹ de tous les autres : « The state in fact obtain (...) is the true description of the world »².

Concernant ce premier découpage ensembliste nous avons donc trois niveaux de description : a) le monde S^3 comme ensemble de tous, b) les états du monde : $s_1, s_2, s_3, \dots, s_n$, et c), l'état du monde qui est le cas et qui met fin à l'ignorance lorsqu'il advient. L'état du monde qui se réalise représente le passage du possible au réel⁴.

À partir du concept d'état du monde Savage construit le concept d'événement, les événements sont notés A, B, C , etc. : « An event is a set of states »⁵. Un événement sera donc un sous-ensemble A – sur lequel pourra être construite une algèbre booléenne – de S ($A \subset S$) ayant un nombre n de s comme éléments ($s \in A$)⁶. Par exemple dans le cas de notre boîte d'œufs, le sujet peut se demander si exactement un seul est pourri, ici il s'agit d'un événement qui a l'ensemble de 12 états du monde comme éléments. Si par contre, il se demande si *au moins* un œuf est pourri, l'événement de référence aura cette fois 2^{12-1} états du monde comme éléments⁷.

Afin de pouvoir appliquer les opérateurs de l'algèbre de Boole, Savage introduit deux événements d'un genre particulier : l'événement universel (S) et l'événement vide (noté O)⁸. Ces deux concepts ont une fonction importante tant au niveau opératoire que conceptuel, en particulier lorsqu'il va s'agir de conditionaliser sur les événements : le fait que l'événement

¹ Le concept d'exclusion est ici central, les états du monde, tout comme les événements sont disjoints.

² Savage (1972), *ibid*.

³ La question de savoir si S est fini ou infini est importante dans le cadre de la théorie de la mesure qui sera élaborée après le quatrième axiome, nous reviendrons sur ce problème à ce moment-là.

⁴ Sur le concept d'« états du monde », voir Fishburn 1970, p 163-174. Son approche est plus générale que celle de Savage mais néanmoins très éclairante.

⁵ Savage (1972), p 10.

⁶ Voir Fishburn (1970), p 130-131, et Kreps (1988), p 116-117 ; voir aussi pour plus de détail, l'article essentiel et fondateur de Kraft/Pratt/Seidenberg (1959), ainsi que Fishburn (1986), p 336 ; et Suppes (1994). Toutes ces considérations sont essentielles dans le cadre de la définition d'une probabilité qualitative telle que définie par de Finetti (1931) et (1937).

⁷ Cet exemple simple est particulièrement pertinent pour comprendre comment Savage découpe, de façon aussi précise que possible, la réalité. En fait le découpage doit être opéré *en fonction* de la décision à prendre. Nous pourrions dire qu'il faut que le sujet soit capable d'*ajuster* le contexte aux décisions qu'il va prendre, et donc par là même d'évacuer les éléments non pertinents. À ce sujet voir Picavet (1996), p 197. Nous reviendrons sur ces problèmes d'ajustement du contexte décisionnel en fonction du système de référence des sujets dans notre troisième chapitre, nous resterons assez proches des vues de Savage.

⁸ Dans le calcul ils vont jouer le rôle du 0 et du 1 en algèbre de Boole ; voir Savage, p 10. L'événement universel a comme éléments tous les états du monde possibles pour une décision donnée, l'événement vide lui n'a aucun état du monde comme élément, il correspond à l'ensemble vide. Sur le rôle opératoire de ces concepts en logique et en théorie des ensembles, voir Halmos (1965) ; Suppes (1972) ; Lepage (2010).

considéré soit non nul est une condition sine qua non de la conditionalisation¹, puisque la probabilité $p|A = p$ si A est nul.

- Le deuxième ensemble que Savage introduit est l'ensemble F des conséquences, notées f, g, h , etc.² : « A consequence is anything that may happen to the person »³.

Ce deuxième ensemble est nécessaire pour formaliser les objets ultimes des préférences que sont les conséquences, mais aussi pour construire de façon originale le concept d'acte. La conséquence est *le résultat commun de l'acte et de l'inscription dans la réalité d'un état du monde donné*, état du monde qui voit son statut se modifier alors en « état de la personne » : « Consequences might appropriately be called states of the person, as opposed to states of the world »⁴.

Afin de rendre plus clair ce qu'il veut démontrer, Savage utilise un exemple bien connu⁵ visant à montrer comment dans un contexte élémentaire et assez épuré se découpe la réalité selon les concepts de monde, états du monde et conséquences, et comment dans cette structure théorique doit être modélisé le concept central d'acte⁶.

L'exemple, que nous allons présenter ici, est le suivant : voulant aider sa compagne – qui a déjà cassé cinq œufs dans un saladier afin de préparer une omelette – un homme s'apprête à casser le sixième œuf qui est posé sur le comptoir de la cuisine. Le protagoniste ne

¹ Voir, Skyrms (1986), p 138-148 ; Baird (1992) p 201-233 ; Hacking (2004) p 57-66.

² Cet ensemble peut être fini ou infini. L'axiome 7 viendra formaliser la finitude de cet ensemble pour les actes. Savage envisage cependant le cas d'actes ayant des conséquences infinies en analysant le pari de Pascal par exemple ; voir Savage (1954) p 77.

³ Savage (1954), p 13. Il y a un certain conséquentialisme de la théorie de Savage ainsi que dans l'axiomatique N/M, qui doit être connecté avec son pragmatisme : les actes sont *en vue* des conséquences, et la délibération qui précède est *en vue* de l'acte. Nous retrouvons ici un schéma proche de celui proposé par Aristote dans l'*Éthique à Nicomaque*, voir par exemple, la première partie du livre III. Il y aurait des parallèles à mener entre le rôle du syllogisme pratique chez Aristote et la théorie de la décision chez Savage.

⁴ Savage (1954), p 14.

⁵ L'exemple de « l'omelette de Savage » est un exemple archétypique en théorie de la décision, il est systématiquement repris. Mentionnons ici les approches de Luce/Raiffa (1957), p 276-277, avec une matrice parfaitement adéquate pour décrire les décisions en contexte d'incertitude ; et l'excellente analyse de Jeffrey (1976), p 361-371.

⁶ Ici, nous observons encore un léger écart avec les théories de Ramsey (1926) et N/M (1947). Savage est en effet le seul des auteurs que nous étudions à considérer le concept d'acte et de préférences sur les actes comme premiers dans sa théorie. Les théories précédemment étudiées parlaient avant tout de préférences sur des différences de valeurs entre des mondes possibles (Ramsey) ou de préférences sur les événements construits comme des loteries, ces deux notions bien qu'étant très proches ne sont toutefois pas identiques ; voir à ce propos Granger (1988), p 290, qui souligne bien que les préférences sur les actes chez Savage sont des préférences sur des fonctions et non plus sur des éléments pris en eux-mêmes. Savage s'inscrit ici, à notre avis, dans une certaine forme de pragmatisme, justement parce qu'il pose le concept d'acte comme premier. À notre sens, cela renforce la visée normative de sa théorie ; voir Baron (1988), p 247 ; les exemples de Gilboa et Al. (2009), p 288 sont très éclairant même s'ils sont critiques ; voir aussi Savage (1968), p 7.

sait pas si cet œuf est frais ou non. Que doit-il faire ? C'est ici qu'intervient le concept d'*acte* que Savage définit ainsi : « An act is a function attaching a consequence to each state of the world »¹.

Revenons à notre exemple. Notre monde S est divisé en deux états : œuf frais/œuf avarié, et j'ai trois actes à ma disposition : casser l'œuf dans le saladier, casser l'œuf dans une soucoupe, jeter l'œuf. C'est la combinaison de ces actes avec les états du monde qui va produire les conséquences (en l'occurrence six ici). Par exemple si l'œuf est frais et que je choisis de le jeter la conséquence sera : une omelette avec cinq œufs et un œuf frais gaspillé. Autre cas de figure : si l'œuf est avarié et que je le mélange aux autres sans inspection la conséquence sera le gaspillage de cinq œufs frais et pas d'omelette. Dans le cas où je choisis d'inspecter l'œuf, qu'il soit frais ou non, la conséquence sera une omelette avec cinq ou six œufs et une soucoupe à nettoyer.

Ce qu'il est important de retenir ici, et à notre sens, c'est un des éléments les plus originaux de la théorie de Savage, c'est la façon dont est construit le concept d'acte : *un acte est une fonction (les actes sont notés f et l'ensemble des actes, qui est donc un ensemble de fonctions est noté F), c'est-à-dire la mise en relation, ou plutôt l'application de l'ensemble des états du monde (S) vers l'ensemble des conséquences (F)*². Plus précisément un acte est une fonction *surjective* des éléments de S vers les éléments de F , avec la possibilité qu'à deux éléments s, t , de S corresponde un seul élément f de F .

Les actes, les états du monde et les conséquences constituent les structures fondamentales de la théorie de Savage³. Il s'agira donc ici d'introduire en *premier lieu* une relation de préférence sur les actes, puis de « *faire passer* » cette relation vers les conséquences (axiome 3) et enfin vers les événements (axiome 4). Ce sera d'ailleurs à partir des préférences sur les événements que vont être définies les probabilités subjectives.

¹ Savage (1972), p 14. L'auteur prend bien soin de noter que deux actes ayant les mêmes conséquences dans tous les états du monde peuvent être considérés comme semblables.

² Ou, comme le souligne Granger (1988) : « Un acte n'est autre qu'une application fonctionnelle de l'ensemble des sous-ensembles d'états, appelés 'événements', dans celui des conséquences », p 290. En fait Granger associe états du monde et événements, ce que ne fait pas Savage. Plus généralement, sur la théorie des fonctions ; voir Halmos (1965), p 40-44 ; Suppes (1972), p 57-90.

³ Voir, Fishburn (1970), p 192, la présentation de la théorie de Savage par cet auteur reste la meilleure qui existe, elle respecte aussi bien la lettre que l'esprit du « maître », peut-être justement en raison du fait que Fishburn a travaillé un certain temps avec Savage ; Kreps (1988), propose une approche qui bien que moins orthodoxe – cet auteur reconstruit parfois la théorie de Savage – est infiniment plus pédagogique, par ailleurs les schémas qu'il utilise sont très pertinents. Du point de vue français, concernant les fondements de l'axiomatique de Savage, on peut lire, Granger (1988), p 290 ; et, Picavet (1996), p 193-194.

Avant de passer à l'axiomatique proprement dite, revenons sur la fin de la section 2.5 qui est pour Savage l'occasion de faire quelques remarques importantes, notamment à propos du rôle normatif¹ de la théorie qu'il est en train d'élaborer : « The point under discussion may be symbolized by the proverb, 'look before you leap', and the one which it is opposed by the proverb, 'you can cross a bridge when you come to it' »².

Il y a un aspect planificateur, voire même « régulateur »³ de la théorie de la décision qui en substance signifie ceci : dans un contexte décisionnel donné le sujet doit pouvoir recenser toutes les possibilités qui s'offrent à lui, ces états du monde possibles regroupés en événements doivent être connectés à des degrés de croyance qui vont eux-mêmes servir à déterminer des probabilités subjectives. La croyance, comme chez Ramsey, constitue ici un élément central. Dans le cadre de l'axiomatique de Savage il s'agit aussi de penser la croyance comme ce qui nous détermine à l'action⁴. Cependant cet aspect planificateur de la théorie devra être limité et circonscrit à des domaines très petits, que Savage baptisera « small worlds »⁵.

C'est dans l'écart que l'on peut constater entre ce que signifient les deux proverbes, que se situe métaphoriquement la théorie de la décision. Dans l'absolu et dans l'optique hautement idéalisée de la théorie de Savage, le sujet devrait être capable d'organiser selon la structure qui vient d'être proposée – monde, états du monde, événements, conséquences – l'ensemble de son existence « décisionnelle »⁶.

2) Les cinq premiers axiomes.

Récapitulons donc les éléments théoriques, dont nous disposons et qui formalisent la réalité décisionnelle du sujet :

¹ Aux yeux de Savage il s'agit là d'une différence entre sa théorie et celle de N/M, voir p 97.

² Savage (1954), p 16. Picavet analyse finement ce passage p 191.

³ Au sens kantien du terme.

⁴ Voir Picavet (1996), p 206.

⁵ Savage analyse ce concept en relation avec sa théorie de l'utilité aux pages 82 à 89. Il s'agit en fait de se demander quelle sont les conditions mathématiques de la délimitation des problèmes décisionnels, ou, comme le dit l'auteur : « what constitute a satisfactory isolated decision situation », p 83. Du point de vue formel il s'agit de travailler sur une partition de S en sous-ensembles (dont certains seront ignorés, selon des critères formels précis) qui constitueront les petits mondes avec leurs éléments ; voir Savage (1954), p 85, et Gilboa (2009).

⁶ Cependant cette thèse est immédiatement limitée par l'exemple que prend Savage : la théorie en construction ne nous permet même pas de planifier un pique-nique tant le nombre d'états du monde possibles à prendre en compte est important.

- Un ensemble S : ensemble de tous les états du monde dans une situation donnée.
- s, t, u , etc. : les états du monde, éléments de S ($s \in S$).
- Un ensemble F : ensemble des conséquences.
- f, g, h, k , etc. : des conséquences particulières ($f \in F$).
- A, B, C , etc. : des événements, qui sont des sous-ensembles de S ($A \subset S$) ayant des états du monde comme éléments ($s \in A$). À partir de ces événements seront construites des algèbres de sous-ensemble ($Coll A$) de S .
- L'ensemble F des fonctions ou applications de S dans F , c'est-à-dire l'ensemble des actes.
- f, g, h , etc. : des actes particuliers ($f \in F$).
- $f(s) = f, s \in A$: l'acte qui donne f si s est dans A et une autre conséquence f' sinon.

Axiome 1 :

« The relation \leq is a simple ordering among acts »¹

Ce premier axiome d'ordonnement sur les actes est exactement du même type que celui de N/M qui lui concernait des loteries, et la relation de préférence a ici les mêmes propriétés² que celles étudiées dans la partie précédente. Savage définit ainsi la relation : $f \leq g$ signifie que l'acte g est *faiblement* préféré à l'acte f ³. Cette relation doit être transitive. Le fait que \leq soit un préordre total implique qu'il n'y a pas d'acte qui soit incommensurable ou qui ne puisse être comparé avec un autre⁴. Par ailleurs, la relation de préférence faible ne doit faire référence à aucun état du monde, elle est donc indépendante des états (« state independant »). Savage déduit de cet axiome le théorème suivant : « If F is a finite set of acts, there exist f and

¹ Savage (1954), p 18.

² Sur lesquelles nous ne reviendrons pas. Nous renvoyons à notre étude du premier axiome chez N/M. Les relations de préférence stricte et d'indifférence sont déduites de façon habituelle ; voir, Fishburn (1970), p 192, et (1986) ; Kreps (1988), p 128.

³ L'auteur explique, sans entrer dans les détails, les raisons de cette définition, plutôt que l'usage d'une définition plus directe. Pour lui, la relation ainsi définie est plus facile à utiliser d'un point de vue technique ; voir Savage (1954), p 17-18.

⁴ Comme nous l'avons vu dans le cadre de la théorie de N/M, cela pose de sérieux problèmes du point de vue descriptif et normatif ; voir, Peterson (2009), p 148.

g on F such that for all g in F , $f \leq g \leq h$ »¹.

C'est un théorème important en ce sens qu'il formalise que, du point de vue de la théorie de la décision, dans une situation décisionnelle donnée il existe un acte qui est faiblement préféré à tous les autres, et dont on peut penser qu'il sera choisi par le sujet². Théorème qui est donc en fin de compte assez en accord avec l'intuition que nous avons du concept d'ensemble d'actes disponibles. Concernant l'axiome I et l'ensemble de la théorie en construction, Savage met en avant deux types d'interprétations : normative et prédictive (empirique). Pour l'auteur, l'aspect prédictif qui nous indique comment les choses pourraient se passer dans la réalité est le moins intéressant³ ; c'est l'interprétation normative qui retient son attention : « Second, it could be regarded as a logic-like criterion of consistency in decision situations »⁴. Savage se situe du point de vue logique et normatif, il nous indique que l'axiome en question peut avoir un rôle correctif, car les sujets commettent parfois des erreurs⁵ qui passent inaperçues à leurs yeux, et que la théorie de la décision permet d'éviter. Une importante partie de l'entreprise de Savage tourne autour cette idée de normativité et de prescription, comme en témoigne cette phrase : « The main use I would make of P1 and its successors is normative »⁶. Les axiomes doivent pouvoir *servir* au sujet afin qu'il puisse améliorer ses décisions si nous le mettons en face de certains types d'incohérences. Par exemple, si le sujet a des préférences incohérentes de type « Dutch Book », nous pourrions le lui faire remarquer et par là même lui donner la *possibilité* de se corriger⁷.

Remarquons que l'exposition de ce premier axiome permet à Savage d'établir un parallèle strict entre la logique et la théorie de la décision du point de vue normatif : si la logique peut et doit être considérée comme la théorie *normative* du raisonnement valide⁸,

¹ Savage (1954), p 18.

² Voir, Picavet (1996), p 211. Ce théorème peut être mis en relation avec l'axiome 5 de non-trivialité.

³ Voir, Savage (1954), p 19. Ce que l'auteur appelle prédictif se rapproche de ce que Baron (1988), p 31-34 appelle la dimension descriptive de la théorie.

⁴ Savage (1954), p 19. Pour l'auteur, il y a un parallèle – abordé à plusieurs reprises dans l'ouvrage – entre la logique et la théorie de la décision ; voir, Picavet (1996), qui revient à de nombreuses reprises sur ce sujet.

⁵ Erreurs de différents types que nous analyserons dans notre prochaine partie, notamment avec la théorie des heuristiques et des biais de Kahneman et Tversky.

⁶ Savage (1954), p 20.

⁷ Comme le remarque de Finetti (1989), il faut garder à l'esprit que ce type d'incohérence peut faire perdre au sujet tout ce qu'il a. Il ne s'agit donc pas d'incohérence purement théorique, l'arnaque est réalisable. La façon dont Savage va réagir au paradoxe d'Allais est à cet égard exemplaire de ce qu'il entend par une utilisation normative des axiomes, nous y reviendrons en conclusion.

⁸ Sur ce sujet la littérature est plus qu'abondante et commence dès *Les premiers analytiques...* Mais on peut se reporter entre autres à Quine (1972), p 11-15 ; Bergmann/Moor/Nelson (1998), p 1-6 ; et, Lepage (2010), p 1-3.

l'axiomatique de Savage devra être considérée comme la théorie *normative* de la décision cohérente en contexte d'incertitude¹.

Axiome 2 :

Le *principe de la chose sûre* (PCS) et P2 qui en est le prolongement permettent d'introduire de façon formelle le concept d'événement et de le relier au concept d'acte par l'intermédiaire de la notion de préférence conditionnelle². Le PCS est une clause d'indépendance forte dont la clé réside dans le concept d'événement³. C'est le principe de la théorie de Savage qui a fait couler le plus d'encre⁴, et c'est sur ce principe que vont se centrer les principales critiques de la TUA⁵ comme le montrent les travaux d'Allais (1951), Ellsberg (1961), Slovic/Lichtenstein (1968), et Kahneman/Tversky (1979)⁶.

Avec le PCS et l'axiome 2, nous assistons, dans le cadre du *mouvement* de constitution de l'axiomatique, à un moment de transition entre l'ordonnancement sur les actes et l'ordonnancement sur les événements qui fera l'objet de l'axiome quatre et qui aura pour fonction d'introduire l'existence d'une probabilité qualitative. Cependant, pour assurer le passage de la relation \leq des actes vers les événements, il faudra d'abord passer des actes vers les conséquences, ce sera l'axiome trois qui permettra cette opération⁷. Le PCS qui est,

¹ Savage va revenir sur ce problème, p 59, son propos est analysé par Picavet (1996), p 222-228 ; et par Gilboa (2009) qui lui se situe dans le cadre d'une évaluation générale du bayésianisme.

² Il est important de remarquer que cette notion quoique présente chez Ramsey sous une forme implicite, en effet le philosophe britannique parle seulement de probabilités conditionnelles, disparaît dans l'axiomatique de N/M.

³ Il y a quelques différences entre le PCS et l'axiome 2 sur lesquelles nous reviendrons, notons toutefois celle-ci immédiatement : dans le PCS il s'agit de mettre en rapport *un* acte avec *un* événement, dans P2 nous avons *quatre* actes et *un* événement. En ce qui concerne le nom « PCS », il y a des auteurs, notamment Kreps (1988), qui préfèrent le réserver à l'axiome 7.

⁴ La plupart des travaux sur Savage, quand ils ne centrent pas leurs analyses de façon critique sur le PCS, lui réservent une part importante. En fait le PCS est souvent introduit avec les critiques qui lui sont adressées ; voir Gilboa et Al. (2009), mais le même auteur dans son ouvrage de 2009 déjà cité prend soin de présenter d'abord les axiomes, p 97-105, et ensuite la critique, p 123-137. C'est la voix que nous emprunterons aussi. On trouve de bonnes et fines analyses du PCS dans Karni (2009), p 5-6 ; Aumann et Al. (2005) : tout l'article porte sur le rapport central entre le PCS et la notion de conditionalisation ; enfin, concernant le principe d'indépendance en théorie de l'utilité, et par définition le PCS, l'article de référence reste celui, très synthétique, de Fishburn/Wakker (1995), p 1136-1138. Cet article, outre le fait qu'il analyse le lien problématique qui existe entre le PCS et l'idée de maximisation, montre aussi comment le PCS peut-être considéré comme l'équivalent de l'axiome « d'indépendance » de N/M, et ce de façon formelle.

⁵ Voir, entre autres, Allais (1953) ; Edwards (1954) ; Gilboa et Al. (2009), p 291-296 ; et Gilboa (2009), p 123-137.

⁶ Pour une analyse d'ensemble des critiques adressées au PCS, voir McClennen (1983). Comme nous le remarquons plus haut à propos de N/M, c'est parce que l'axiome d'indépendance est ce qui permet d'introduire l'hypothèse de maximisation de l'utilité attendue qu'il a été la cible de toutes les critiques : cet axiome est le cœur de la TUA. Voir Fishburn/Wakker (1995), p 1130-1132 ; Shafer (1986), p 475. Voir aussi la critique de Amihud (1979).

⁷ Il est important de bien comprendre, même si ce n'est pas toujours évident en raison de la difficulté du texte, qu'il y a un *mouvement* dans la constitution de l'axiomatique de Savage, un mouvement qui va des préférences faibles sur les actes vers

comme le souligne Savage, un principe extralogique (« extralogical »)¹ s'énonce ainsi dans sa forme générale : « If the person would not prefer f to g , either knowing that the event B obtained, or knowing that the event $\neg B$ obtained, then he does not prefer f to g »². Formellement, pour tout $f, g \in F$ et pour tout $B \subset S$,

$$\{[(f \geq g) | B] \wedge [(f \geq g) | \neg B]\} \supset (f \geq g)$$

Où, selon la formulation suivante que propose Savage, avec la relation de préférence stricte, pour tout $f, g \in F$ et pour tout $B \subset S$,

$$\{[(g > f) | B] \wedge \neg[(f > g) | \neg B]\} \supset (g > f)^3.$$

Le PCS exprime l'idée suivante, qui sera affinée avec P2⁴ : un événement B , dont le sujet *ne sait pas* si lui ou son complémentaire $\neg B$ va se réaliser, ne peut avoir aucune influence sur l'ordonnancement des préférences entre deux actes f et g . Autrement dit, les préférences sur les actes sont *indépendantes* des événements pour lesquels les conséquences sont équivalentes, puisque si B advient je préfère g et si $\neg B$ advient je préfère aussi g . Comme le remarque Savage, le PCS ne doit pas être, à strictement parler, inclus dans l'axiomatique comme une clause formelle, car il fait intervenir des concepts que la théorie de la décision ne peut formaliser, comme la connaissance ou la possibilité⁵. Il faut noter que dans l'esprit de

les probabilités subjectives, et des probabilités vers l'utilité. Seuls quelques auteurs comme Luce/Raiffa (1957), p 302 ; Granger (1988), p 290-291 ; et Picavet (1996), p 215, mentionnent ou analysent ce mouvement qui va en fait des axiomes 1 à 4, les autres axiomes étant des clauses plus techniques. Nous prendrons soin de bien marquer les moments de cette dynamique.

¹ Savage (1954), p 21. Ce qui nous invite immédiatement à nous poser cette question : si le PCS n'est pas un principe logique, un axiome, dans quelle mesure peut-il être considéré comme normatif ? Dans quelle mesure un agent n'obéissant pas au PCS peut-il être considéré comme irrationnel ? Voir, Ellsberg (1961) ; Picavet, 1996, p 199-200 ; et Aumann et Al. (2005), p 303.

² Savage, *ibid.*

³ Il est bien évident que dans les deux cas l'événement de référence ne doit pas être impossible ou nul.

⁴ Remarquons que parfois le PCS et l'axiome 2 sont simplement confondus ou mal distingués. En fait il faut les distinguer pour la raison que l'axiome 2 ne peut exprimer tout ce que « formalise » le PCS, notamment le rapport aux concepts de connaissance et de possibilité. Nous pourrions dire que l'axiome 2, plus restrictif, sous-détermine le PCS qui a des implications philosophiques échappant à la formalisation. Il faut garder cette distinction à l'esprit, car elle reflète la finesse des distinctions conceptuelles opérées par Savage.

⁵ Voir, Savage (1954), p 22. Il y a un aspect épistémique et doxastique du PCS. En fait il y a un aspect épistémique qui traverse l'ensemble du bayésianisme, et qui est lié en partie aux problèmes de conditionalisation ; voir entre autres Stalnaker (1970) ; Lewis (1976) ; Oaksford/Chater (2007) et Chater/Oaksford (2008) ; Lepage/Morgan (2011). Savage reviendra largement sur ce problème aux pages 43-55, où il s'agira justement d'analyser comment le sujet peut réviser ses croyances, en partie grâce à l'utilisation du théorème de Bayes. En fait il s'agit d'analyser le processus d'apprentissage par

Savage, le PCS va de soi : « I know no other extralogical principle governing decisions that finds such ready acceptance »¹. L'exemple introductif qu'utilise l'auteur est d'ailleurs assez simple : un homme d'affaires envisage d'acheter une parcelle de terrain et trouve le résultat de la prochaine élection présidentielle pertinente pour la réalisation de son achat². Plus précisément, il se demande à lui-même si le fait de *savoir* que le prochain président soit républicain (événement B) serait déterminant, de façon favorable, en vue de son achat (acte g , l'acte f étant ici de ne pas acheter). Parallèlement il se pose la même question pour $\neg B$ (le président élu est démocrate) et il se dit que le *savoir* de $\neg B$ serait lui aussi pertinent en faveur de g ³. Il en conclut donc ceci : que B ou $\neg B$ advienne, je préfère g à f ⁴. Comme le remarque l'auteur, le PCS est contraignant dans les rares cas où il peut être appliqué.

Il est temps maintenant de se demander quel rôle joue le PCS dans l'économie générale de la théorie que Savage entend élaborer. Tout d'abord, le PCS permet d'introduire la notion d'indépendance, qui, dans l'esprit des TUA « néo bernoulliennes »⁵, garantit l'existence du phénomène de maximisation de l'utilité attendue⁶. Cette notion sera formalisée par l'axiome 2. Ensuite, le PCS permet à l'auteur d'introduire deux autres notions centrales : la relation de *préférence conditionnelle* et celle d'*acte modifié*. Concernant la première notion, la question est la suivante : quel sens attacher au fait que $f \geq g$ si l'on sait que B va arriver⁷ ? Comme le remarque Savage : « Under any reasonable interpretation, the matter would seem not depend on the values f and g assume at state outside B »⁸. Le seul élément qui importe ici ce sont les valeurs que vont prendre les conséquences des actes si B est le cas. Ces valeurs sont

expérience. À ce sujet on peut se reporter à Oaksford/Chater (2007), p 67-99, Hacking (2004), p 185-198, Skyrms (1986), p 189-194.

¹ Savage (1954), p 21.

² L'exemple est présenté à la page 21.

³ Si nous y regardons de plus près, c'est en fait un principe assez proche de l'intuition que nous pouvons avoir dans certaines situations, et c'est probablement ce que Savage entend formaliser. Il n'est pas rare en effet que dans nos existences nous nous trouvions confrontés à des situations décisionnelles du type : que A ou non A se produise, je vais faire ceci. Il est bien clair que nous parlons ici de situations banales qui n'impliquent aucune conséquence « dramatique » pour le sujet ou son entourage, car dans ce cas-là les choses sont bien entendu différentes. Mais les options ou les événements ultimes sont – heureusement – rarement convoquées dans notre environnement décisionnel

⁴ Pour un raffinement critique de cet exemple, voir Aumann et Al. (2005).

⁵ C'est ainsi qu'Allais (1953) nomme les théories de N/M (1947), Samuelson (1952) et Savage (1951).

⁶ C'est justement sur cette idée que portent les critiques d'Allais et sur le fait que d'une part les axiomes ne sont absolument pas normatifs, d'autre part que les sujets ne maximisent pas l'utilité attendue. Cette critique sera reprise, cette fois du point de vue de la psychologie, par Edwards (1954).

⁷ Il faut noter au passage que l'opération de conditionalisation temporalise de façon nécessaire le processus décisionnel, l'événement sur lequel nous conditionnons se situe toujours dans le futur, même si il s'agit d'un futur proche comme dans la théorie de N/M.

⁸ Savage (1954), p 22.

représentées comme suit : $f(s)$, $g(s)$, etc. Ce qui symbolise la valeur de l'acte f ou de l'acte g si $s \in B$ ¹. Cependant nous pouvons faire l'hypothèse que dans tous les cas étrangers à B , c'est-à-dire dans $\neg B$, les deux actes s'accordent : $f(s) = g(s)$. Comment cette opération est possible ? En fait il s'agira de *modifier* les actes de façon à faire coïncider leurs conséquences en dehors de l'événement de référence. Une fois cette modification opérée, nous avons d'une part dans B , $f(s) \neq g(s)$ ² mais $g(s) = g'(s)$ et $f(s) = f'(s)$ ³, d'autre part, dans $\neg B$, $f(s) = g(s)$ et $f'(s) = g'(s)$. Ce qui va nous permettre de définir une relation de *préférence conditionnelle* en fonction de B :

« If after being modified so as to agree with one another outside of B , f is not preferred to g ; then f would not be preferred to g , if B were known. The notion will be exposed formally by saying that $f \leq g$ given B »⁴.

C'est à partir du PCS et de cette définition que Savage va introduire son deuxième postulat :

- « If f, g and f', g' are such that
- 1) In $\neg B$, f agrees with g , and f' agrees with g'
 - 2) In B , f agrees with f' , and g agrees with g'
 - 3) $f \leq g$
 - 4) Then $f' \leq g'$ »⁵

Comme nous le voyons, P2 est un axiome d'indépendance qui signifie ceci : les préférences sur les actes ne doivent pas être affectées par les événements à l'intérieur desquels

¹ Il existe différents types de notations pour représenter les actes et leurs valeurs conditionnelles aux événements. Par exemple, Karni (2005), p 7, utilise la notation suivante, très claire : soit un événement B et deux actes f et g , $f_E h$ représente un acte tel que $(f_E h)(s) = f(s)$ si $s \in B$ et $(f_E h)(s) = h(s)$ autrement. Cette formalisation a le mérite d'être très synthétique et elle se rapproche de celle utilisée par Peterson. Gilboa (2009), p 99, opte pour un formalisme plus chargé, mais qui a l'avantage d'explicitement toutes les composantes d'une alternative : pour un événement B , deux actes f et g et un état du monde s : $f_B^g(s) \{ \begin{matrix} g(s) & s \in B \\ f(s) & s \in B^c \end{matrix} \}$. Picavet (1996), utilise un formalisme encore différent, à la fois très simple et explicite, il se rapproche de celui de Karni. D'une façon plus simple, pour un état du monde s , et un événement B , l'acte f peut être formalisé ainsi : $[f(s) \text{ si } s \in B; g(s) \text{ si } s \notin B]$, comme la loterie qui garantit la rétribution $f(s)$ si s est dans B et $g(s)$ autrement. Voir, Picavet (1996), p 198.

² Différence importante puisque c'est elle qui permet d'établir la préférence.

³ Les actes f' et g' sont les actes f et g modifiés de façon arbitraire, à cette condition qu'ils s'accordent dans $\neg B$. Voir Savage (1954), p 22 ; Picavet (1996), p 198 ; et Gilboa (2009), p 98.

⁴ Savage (1954), ibid.

⁵ Savage (1954), p 23.

les actes ont des résultats, des conséquences identiques¹. Autrement dit, dans le cadre de l'axiome 2, il faut considérer les actes f, g, f', g' , conditionnellement à B puisque dans $\neg B$ ils sont identiques, d'autre part, les actes modifiés ne le sont que dans $\neg B$. Ce qui permet de montrer à la fois que la notion de préférence conditionnelle est un ordre simple² et qu'elle est bien définie³. Les préférences sur les actes doivent dépendre de la valeur de ceux-ci seulement lorsque leurs conséquences sont différentes⁴. L'intention de Savage lorsqu'il élabore cet axiome, outre l'introduction de la propriété d'indépendance qui est fondamentale, c'est de pouvoir « faire passer » la relation \leq des actes simples vers les actes conditionnés sur des événements de référence, et ce en vue d'importer la relation \leq vers des actes conditionnés dits « constants ».

Axiome 3 :

Avec cet axiome nous assistons donc au deuxième mouvement de déplacement de la relation de préférence faible. En effet, le PCS et P2 permettent d'introduire la notion de préférence conditionnelle sur les actes par rapport à un événement de référence. Avec l'axiome 3 nous allons maintenant passer des actes aux conséquences, et ce en vue de pouvoir introduire l'axiome 4, dans lequel les conséquences joueront un rôle important. Cette démonstration va être pour Savage l'occasion d'affiner et de complexifier la relation entre acte et conséquence⁵. Le rôle de P3 sera de montrer comment les préférences entre les conséquences – identifiées aux actes constants qui les produisent – doivent être indépendantes des événements⁶ de façon à ce que : « Knowledge of an event cannot establish a new

¹ Propriété d'indépendance qui est cette fois totalement explicitée, ce qui n'était pas le cas chez N/M, pour lesquels il fallait montrer, que les clauses $3 : B : a$ et $B : B : b$ impliquaient la propriété d'indépendance, ce qu'à brillamment démontré Malinvaud (1952). Voir, Shafer (1986), p 475 ; Gärdenfors/Sahlin (1988), introduction ; Fishburn/Wakker (1995), p 1136-1137.

² « Each of the relations ' \leq given B ' is now easily seen to be a simple ordering », Savage (1954), p 23. Cela lui permet du même coup de définir les relations de préférence stricte et d'indifférence conditionnelles.

³ La définition de la préférence conditionnelle aux événements permet aussi à Savage de donner une définition formelle de l'événement nul. Un événement B pourra être dit nul si pour tout acte $f, g : f \doteq g \mid B$. Le théorème 1 explicite diverses propriétés de l'événement nul ; voir, Savage (1954), p 24 ; Gilboa (2009), p 99-100.

⁴ Pour une représentation graphique tout à fait pertinente nous renvoyons à Kreps (1988), p 129.

⁵ Les deux éléments sont objets des préférences, mais les conséquences sont les objets « ultimes » des préférences. Voir, Picavet (1996), p 203.

⁶ Nous parlons ici de « monotonie par rapport aux événements ». Voir Granger (1988), p 291 ; et, Barberà/Hammond/Seidl (2004), p 729.

preference among consequences or reverse an old one »¹. De façon formelle l'axiome est introduit comme suit :

« If $f \equiv g$, $f' \equiv g'$, and B is not nul ; then $f \leq f'$ given B , if and only if $g \leq g'$ »².

L'aspect problématique de cet axiome c'est l'identité $f' \equiv g'$. Demandons-nous de quelle façon un acte peut-il être rendu équivalent ou identique à une conséquence ? Cette identité ne pourra être réalisée que *si et seulement si* l'acte est constant (« constant act »), c'est-à-dire qu'il donne la même conséquence quelque soit l'état du monde considéré³, ou de façon formelle :

$$(\forall s)[s \in S \supset (f(s) = x)]$$

Cette définition permet de formaliser la relation de préférences entre les conséquences en termes de préférence entre les actes. Une conséquence g sera préférée à une conséquence g' si l'acte constant f auquel on peut l'identifier est préféré à l'acte constant f' , donc :

$$(g \leq g') \equiv (f \leq f')$$

Cependant, le fait que ce type de relation puisse exister ne doit pas nous induire à *confondre* purement et simplement, dans un problème décisionnel donné, les actes avec les conséquences qui leur sont associées. En effet, il ne faut pas oublier que dans le système de Savage, *les actes et les conséquences doivent demeurer deux entités théoriques distinctes* ; et ce, pour la simple raison que bien qu'étant l'objet de préférences, les actes et les conséquences ne font pas l'objet des *mêmes* préférences. Les conséquences sont les entités ultimes des processus décisionnels, les actes sont préférés *en vue* des conséquences.

¹ Savage (1954), p 26. Comme le remarque l'auteur, la connaissance que B sera le cas ou pas ne peut pas non plus transformer la relation de préférence faible en relation d'indifférence. Il semblerait que d'un point de vue descriptif tout ce développement soit assez contre intuitif, c'est ce que souligne Kreps (1988), p 34-38. En effet, il n'est pas rare que le fait de savoir qu'un événement de référence va se produire ou non, altère nos préférences originelles.

² Savage (1954), ibid.

³ Sur ce concept, très utile d'un point de vue opératoire, mais néanmoins problématique, car nous parlons ici d'une classe d'actes très rares dans la réalité décisionnelle, voir Picavet (1996), p 202 ; Peterson (2009), p 149 ; et, Gilboa (2009), p 100.

Ce que montre aussi P3 c'est que le sujet préfère l'acte qui peut le conduire à la conséquence ayant à ses yeux la valeur la plus grande, et ce en faisant abstraction des événements qui pourraient ou non advenir. Nous pouvons dire que nous restons dans l'esprit du PCS, qui mettait les préférences sur les actes « à l'abri » de l'influence de la réalisation ou pas d'un événement de référence, l'axiome 3 opère la même chose, mais cette fois pour les conséquences. Savage se pose d'ailleurs, à ce sujet cette question : « Suppose that $f \equiv g$, $f' \equiv g'$ and, $g \leq g'$, is it reasonable to admit that for some B , $f > f'$ given B ? »¹. Autrement dit, dans quelle mesure la connaissance qu'un événement B donné sera le cas peut-elle renverser nos préférences originelles²? Comme le remarque l'auteur, la réponse à cette question tient à la façon dont nous *interprétons* les concepts d'actes et de conséquences.

Savage va avoir recours à un exemple³ pour clarifier sa pensée. Une personne est invitée à un pique-nique avec des amis et doit se décider à acheter un maillot de bain ou une raquette de tennis (pas les deux, car il manque d'argent). L'événement de référence est le suivant : le pique-nique a lieu au bord de l'eau (B) ou pas ($\neg B$). Comment doit-on considérer les éléments « posséder une raquette » et « posséder un maillot de bain »? Conséquences ou actes?

Si nous considérons ces entités comme des conséquences – de l'acte d'acheter un maillot ou d'acheter une raquette – et si le sujet préfère la conséquence « avoir un maillot » mais que $\neg B$ est le cas, alors sa préférence va se trouver renversée⁴ : le sujet préférera avoir une raquette de tennis. Or, comme le note Savage : « Under the interpretation of 'act' and 'consequence' I am trying to formulate, this not the correct analysis of the situation »⁵. En effet, dans le cadre de la théorie que nous étudions, les éléments « avoir un maillot » et « avoir une raquette » doivent être considérés comme des actes (f et f'), ayant par exemple pour conséquence, si B est le cas, et que le sujet ai choisi f , une séance de natation avec ses amis. C'est le seul moyen pour Savage de « préserver » les préférences sur les conséquences de l'influence d'un événement de référence⁶, car l'acte constant auquel est attaché la conséquence

¹ Savage (1954), p 25.

² L'étude du phénomène du renversement des préférences fera l'objet des études séminales de Lichtenstein et Slovic (1968), mais surtout (1971), nous reviendrons sur ces problèmes dans notre prochain chapitre.

³ Savage (1954), *ibid.*

⁴ Sa préférence deviendra par là même « state dependent », ce que Savage ne veut pas.

⁵ Savage (1954), *ibid.*

⁶ Donc, de les maintenir totalement « étanches » à l'influence des événements.

la détermine dans tous les états du monde, et cet acte (ou plutôt ces actes, car il y en a deux ici) est indépendant de la réalisation de B ou $\neg B$. Autrement dit, quel que soit l'événement, j'opterai pour l'acte qui me conduira, de façon constante vers la conséquence que je préfère.

Axiomes 4 et 5 :

Le quatrième axiome introduit par Savage a pour but de permettre de découvrir les croyances du sujet relatives aux probabilités de la réalisation de deux événements A et B , en relation – *mais non pas en fonction* – avec ses préférences sur deux actes constants identifiés à leurs conséquences f et f^1 . Il s'agit ici du troisième et dernier moment² de l'importation de la relation de préférence qui a pour but de montrer que \leq est un *préordre total sur les événements*, et que cet ordonnancement implique l'existence d'une probabilité subjective qui permettra avec les deux derniers axiomes de déduire une fonction d'utilité.

La question que se pose Savage au début du chapitre 3 : « Personal probability »³ est de savoir comment puis-je découvrir les probabilités qu'un sujet attribue aux événements à partir de son *comportement* dans un contexte de pari. C'est l'introduction des probabilités dans l'axiomatique, et à la différence de N/M, ces probabilités ne sont pas données mais construites par le sujet en fonction de ses croyances. Nous sommes ici encore une fois très proches de Ramsey⁴, et ce pour deux raisons : l'utilisation du concept de pari et l'élimination de l'introspection⁵ ou de l'utilisation du questionnaire comme méthodes révélatrices des préférences sur les événements, en effet : « In the first place, many doubts that the concept 'more probable than' is an intuitive one, open to no ambiguity and yet admitting no further

¹ D'où l'importance de P3, et du rapport entre actes et conséquences. Avec P4 nous nous situons au niveau des croyances, comme le remarque Picavet (1996), p 206, la croyance sert à expliquer l'action. Mais il faut ajouter que la croyance explique l'action, non pas comme, par exemple, une cause pourrait expliquer un effet qui dépend d'elle. Il s'agit ici d'une explication « rétroactive ». Ce sont les préférences sur les actes qui révèlent les croyances, et ces dernières, en retour, éclairent la raison de l'acte.

² C'est d'ailleurs pour cela que P1-P4 forment un tout articulé qui peut être séparé des autres axiomes pour les raisons de l'exposition, les clauses P5-P7 sont des clauses techniques, structurelles. Le rôle normatif de P5-P7 est bien moindre que celui du « bloc » P1-P4 qui est encore le centre de beaucoup de débats ; voir, Shafer (1986) ; Gilboa (2009), p 123. P6 et P7 auront pour fonction de permettre de dériver une fonction d'utilité qui est limitée, nous y reviendrons.

³ Il couvre les pages 27 à 55.

⁴ Avec cette différence notable que Ramsey déduisait les probabilités de l'utilité que le sujet attribue aux états du monde via la mesure de leurs différences de valeurs, alors que Savage ira dans le sens inverse et déduira l'utilité à partir des probabilités personnelles ; voir Ramsey (1926), p 173-176 .

⁵ Remarquons que N/M éliminaient aussi cette méthode.

analysis »¹, et même en admettant que ce concept soit intuitif, l'interrogation directe ne nous renseignerait que sur ce que dit le sujet, non sur la façon dont il se *comporterait* face à une décision en contexte d'incertitude².

C'est donc le pari qui va, comme chez Ramsey, servir de révélateur. Soit l'exemple suivant : un sujet vient de prendre deux œufs dans le réfrigérateur, il y en a un blanc et un marron et il les tient tous deux dans sa main. Nous nous posons la question suivante : « We wonder whether he thinks it more probable that the brown one is good than that the white one is »³. Imaginons que nous proposons 1\$ au sujet si il découvre l'œuf qui est bon, et que sinon nous remplaçons l'œuf pourri par un œuf frais. Dans ces conditions il semble assez naturel de poser ceci : si le sujet est prêt à miser son dollar sur l'œuf marron, c'est qu'il pense qu'il est *plus probable* que celui-ci soit frais⁴. C'est cette intuition que va venir formaliser l'axiome 4⁵ : si,

$f, f', g, g'; A, B; f_A, f_B, g_A, g_B$ are such that :

$$1. \quad f' < f \quad g' < g$$

$$2a. \quad \begin{aligned} f_A(s) &= f, & g_A(s) &= g & \text{for } s \in A, \\ f_A(s) &= f', & g_A(s) &= g' & \text{for } s \in \neg A; \end{aligned}$$

$$2b. \quad \begin{aligned} f_B(s) &= f, & g_B(s) &= g & \text{for } s \in B, \\ f_B(s) &= f', & g_B(s) &= g' & \text{for } s \in \neg B; \end{aligned}$$

$$3. \quad f_A \leq f_B;$$

$$\text{then} \quad g_A \leq g_B.$$

¹ Savage (1954), p 27. Tout ce passage, très important, sur la méthodologie du questionnaire scientifique – tel qu'il sera pratiqué en psychologie cognitive – et que nous discuterons dans notre prochaine partie, est très bien analysé par Picavet (1996), p 206-210 ; voir aussi Andler (2004), p 315-405. À ce propos, les remarques méthodologiques séminales d'Edwards (1954) et de Simmons (1959) sont aussi très importantes.

² Cet aspect de la théorie de Savage est intéressant. Pour l'auteur, la finalité des statistiques est éminemment *pragmatique* comme le souligne par exemple cette phrase à propos de la théorie statistique de la décision qui doit avoir pour but « what action to decide under such circumstances », Savage (1954), p 2, p 154-157 et 159.

³ Savage (1954), p 28. Notons qu'il s'agit ici en quelque sorte de « deviner » quelle probabilité subjective l'agent attribue à l'événement en question.

⁴ Savage fait remarquer qu'une somme trop petite pourrait rendre le sujet indifférent à la question. D'autre part, concernant l'œuf frais de « remplacement » il s'agit ici d'un lot de consolation. Ce concept est important pour la compréhension de l'axiome 4 et aussi pour comprendre la propriété 2 page 32 dans la définition de la probabilité qualitative.

⁵ Savage (1954), p 31.

Cet axiome formalise ceci : pour une conséquence f strictement préférée à f' , et une conséquence g strictement préférée à g' , si les rétributions attachées aux actes constants f_A si s est dans A , f_B si s est dans B d'une part, et g_A si s est dans A , g_B si s est dans B d'autre part, sont égales, à savoir respectivement f et g ; et si par ailleurs le sujet préfère faiblement l'acte f_B à l'acte f_A cela implique qu'il doit préférer l'acte g_B à l'acte g_A .

Ce qui signifie de façon plus profonde que je préfère l'événement B à l'événement A parce que je le trouve plus probable¹. Le ressort de cet axiome c'est le fait que les conséquences sous A et sous B soient équivalentes. En effet les conséquences de f_A si s est dans A et de f_B si s est dans B sont égales, donc si je préfère f_B à f_A c'est que je pense que l'événement de référence B est plus probable que l'événement de référence A , qu'il a donc plus de chances de m'apporter la conséquence f que je préfère originellement à f' . Si les prix attachés aux actes conditionnés par les événements A et B sont les mêmes mais que j'en choisis un plutôt que l'autre c'est que j'évalue l'événement qui conditionne l'acte choisi comme *plus probable* que l'autre, c'est-à-dire comme ayant plus de chance de m'octroyer la conséquence désirée. Cela permet à Savage de montrer que \leq est un préordre total sur les événements².

Au sujet de cet axiome, une remarque s'impose toutefois : cette clause établit une indépendance des événements avec les conséquences (car les préférences sur les événements ne sont pas dues aux conséquences, qui sont identiques pour A et B , mais aux croyances du sujet), seulement cette indépendance ne peut être obtenue que si la première partie de l'axiome (1.) porte néanmoins *explicitement* sur deux préférences entre des conséquences : $f' < f$ et $g' < g$.

P5 qui est une clause technique nous assure qu'il existe au moins une préférence stricte sur une paire de conséquences : « There is at least one pair of consequences f , f' such that

¹ La relation de préférence faible se trouve ainsi déplacée des conséquences aux événements.

² Voir, Gilboa (2009), p 4 ; Shafer (1986), p 468 ; Karni (2005), p 6 ; Picavet (1996), p 193, comme le remarque cet auteur, la relation de préférence faible sur les événements « est fondée sur les croyances », c'est aussi dans ce sens que va l'analyse de Karni. Du point de vue de la clarté de l'exposition et de la pédagogie, c'est encore Kreps (1988), p 130-131 qui est le meilleur. Nous trouvons aussi de très bons exemples dans Ellsberg (1961), p 249.

$f' < f$ »¹. Cet axiome permet d'assurer que l'on peut établir les deux préférences originelles entre f et f' , et entre g et g' . Sans lui, nous pourrions avoir à faire au cas où toutes les conséquences soient équivalentes, ce qui empêcherait tout simplement P4 de fonctionner, puisque comme nous venons de le voir, la première condition est essentielle. P5 est donc un axiome de non-trivialité.

À l'issue de ce développement Savage est parvenu à introduire une probabilité qualitative sur les événements notée \leq^* et qui doit être lue ainsi : $A \leq^* B$ signifie que B est faiblement plus probable que A *du point de vue subjectif*. C'est-à-dire que le sujet *croit* que B est plus probable que A . Cette relation doit satisfaire les propriétés suivantes :

1. \leq^* est un préordre total sur les événements.
2. $B \leq^* C$ si et seulement si $B \cup D \leq^* C \cup D$, pour $B \cap D = 0$ et $C \cap D = 0$. Il s'agit d'une forme de principe d'indépendance concernant les probabilités subjectives, qui signifie que \leq^* sur B et C est indépendante de la probabilité d'obtenir un lot de consolation D dans le cas où ni B ni C ne serait le cas.
3. $\emptyset \leq^* B$, $\emptyset <^* S$ ²

À ces conditions et à la condition de non-trivialité que formalise P5, qui comme nous venons de le voir est une clause qui va de soi, \leq^* est une relation de *probabilité qualitative*³ attachée aux événements. L'introduction de ce type de probabilité clôt le premier moment de l'axiomatique, ce faisant, elle ouvre un vaste problème et de vastes perspectives sur la méthode à employer afin de passer d'une probabilité qualitative à une mesure probabiliste quantitative.

3) Les probabilités subjectives et l'utilité.

Le cheminement de Savage a donc été le suivant : passer des préférences sur les actes aux préférences sur les événements et de là établir grâce à P4 l'existence d'une probabilité

¹ Savage (1954), p 31.

² Voir, Fishburn (1986), p 336 ; Suppes (1994), p 19-20.

³ Au sens que donne de Finetti (1937) à ce terme, p 6.

subjective sur ces derniers. Il va s'agir avec P6 et P7 de passer des probabilités subjectives à la mesure de l'utilité attendue d'une perspective aux conséquences finies, pensée sur le modèle d'un pari (gamble)¹. Cependant, avant toutes choses, il est nécessaire à notre auteur d'approfondir sa théorie des probabilités subjectives dans deux directions : tout d'abord il faut passer d'une *probabilité qualitative* à une *mesure de probabilité quantitative*, ensuite, il faut expliciter les mathématiques du concept de conditionalisation sur les événements².

Nous disposons donc avec P4 d'une probabilité qualitative sur les événements notée \leq^* qui a les caractéristiques que nous avons présentées plus haut, caractéristiques qui avaient été établies par de Finetti dans son article de 1937³. Il va s'agir maintenant de faire correspondre cette probabilité qualitative avec une mesure de probabilité quantitative satisfaisant les axiomes de Kolmogorov (1933) :

« A probability measure on a set S is a function $P(B)$ attaching to each $B \subseteq C$ a real number such that :

1. $P(B) \geq 0$ for every B ,
2. If $B \cap C = \emptyset$, $P(B \cup C) = P(B) + P(C)$,
3. $P(S) = 1$ »⁴.

La question que nous pouvons nous poser est celle-ci : pourquoi cette opération – passage du qualitatif au quantitatif – est-elle nécessaire ? Cette opération est rendue nécessaire pour produire une véritable *mesure*, car comme nous l'avons déjà vu chez Ramsey et N/M, la

¹ Savage (1954), p 70-82.

² Nous ne rentrerons pas dans cette démonstration qui va des pages 43 à 46. Seulement, il faut noter que l'utilisation des probabilités conditionnelles est nécessaire dans le cadre de la théorie de Savage et du bayésianisme en général pour rendre compte de l'évolution et de la mise à jour des probabilités subjectives par rapport aux événements (bayesian updating). Sans cette opération, il serait impossible de rendre compte de la dynamique des croyances du sujet. Et, il est dans l'essence même de la croyance de se transformer et d'évoluer, de changer, c'est ce qu'a bien vu toute une part de la philosophie occidentale, à commencer par Platon, voir *République*, V et VI, et plus généralement Pascal Engel, "Les croyances", in D. Kambouchner (Ed.) *Notions de philosophie*, Paris, Gallimard, 1995, tome II, 1-101. Les croyances sont aussi ce qui constitue une bonne partie du tissu de nos « états » intellectuels à propos du monde en général. Il est donc nécessaire d'avoir un outil qui permette de les formaliser. Mais la question de savoir si les probabilités sont un outil pertinent reste centrale dans le débat contemporain comme en témoignent les travaux de Stalnaker (1970), Lewis (1976), et Lepage/Morgan (2011). Voir aussi, pour une vue d'ensemble à la fois générale et précise : Hacking (2004), Gilboa et Al. (2004) ; Oaksford/Chater (2007) et, pour une étude profonde en français, du « style » bayésien, voir Granger (1988).

³ Les axiomes de de Finetti ont donné lieu à bon nombre d'articles et d'analyses, mentionnons les plus pertinents : M.C. Galavotti (1989) et (1991) ; Fishburn (1986) ; Suppes (1994) ; le livre de Gilboa (2009) contient une très bonne analyse de la théorie de l'utilité et des axiomes de de Finetti.

⁴ Savage (1954), p 33.

théorie de l'utilité est une théorie de la mesure¹, or, sans quantification il n'y a pas de mesure possible².

Parmi les auteurs que nous avons étudiés jusqu'à maintenant, Savage est celui qui est allé le plus loin dans l'explicitation mathématique des conditions de possibilité de la mesure. Aussi allons-nous essayer de le suivre d'assez près dans son cheminement. Quelles sont les conditions mathématiques *nécessaires et suffisantes* pour passer d'une probabilité qualitative à une mesure de probabilité quantitative ?

Pour parvenir à déterminer ces éléments il nous faut faire deux choses. D'une part, expliciter les conditions de possibilité d'une mesure en général, ce que Savage ne fait pas ou seulement bien trop allusivement³, et qui nous seront utiles justement pour comprendre avec quels outils formels Savage construit sa démonstration, d'autre part, il faut revenir à la définition que donne l'auteur du concept d'événement : un événement B est un sous-ensemble de l'ensemble universel S^4 . Pour qu'une mesure soit possible, c'est-à-dire pour que \leq^* puisse être exprimée en terme de probabilité quantitative nous allons devoir construire une algèbre booléenne à partir d'une collection ($CollA$) de sous-ensembles (A, B) de S^5 . C'est ici que va entrer en ligne de compte la question de savoir si S est infini ou non. Si S est fini, une algèbre booléenne devra satisfaire les conditions suivantes :

1) $S \in CollA$, (S est un élément de la collection),

¹ La « déduction » du théorème de la représentation et de son unicité font tous deux appels à des séries d'opérations mathématiques, dont les étapes se rapprochent d'assez près de ce que nous avons vu plus en détail chez N/M. Pour une approche plus exhaustive de ces problématiques, voir Suppes/Zinnes (1963) ; Krantz/Luce/Suppes/Tversky (1989), Vol II ; et, Dicks/Tournois/Flieller/Kop (1994), p 15-60.

² Voir, Carnap (1960), qui explore ce problème dans la deuxième partie de son ouvrage, et qui applique une méthode assez proche de celle que nous avons vu chez Ramsey et N/M – et qui sera aussi celle de Savage – aux concepts de temps et de longueur, p 78 à 86. La discussion que propose Carnap aux pages 62-69 est très importante concernant les fondements de la physique et des sciences de la mesure en général, et il est très intéressant de la lire en parallèle avec les considérations méthodologiques et épistémologiques de N/M (1944), p 20-25. Concernant le problème général de la mesure, nous renvoyons à la profonde analyse de Granger (1988), p 256-258.

³ Par exemple aux pages 40-43.

⁴ Ce qui signifie que pour un nombre n d'événements, l'ensemble des sous-ensembles de S est égal à 2^n . Détail important puisque c'est sur ce nombre cardinal que sera calculée la probabilité de l'événement de référence. Sur la nécessité de construire une mesure d'après un ensemble, voir Granger (1988), p 257.

⁵ Ici, nous sortons un peu de la théorie de Savage proprement dite en présentant les prérequis mathématiques généraux de la théorie de la mesure. Nous nous inspirons dans tout ce qui va suivre des travaux de Fishburn (1970), p 129-134, et (1973), p 68-76, (1986) ; Kreps (1988), p 115-126, l'exposé de cet auteur est un des plus pédagogiques qui soit ; Suppes (1994) ; enfin les références majeures pour les problèmes de théorie de la mesure et de déduction du théorème de la représentation sont Suppes/Zinnes (1963) ; et Krantz/Luce/Suppes/Tversky (1971), Vol I et II. Savage fait quelques remarques « en passant » sur ces problèmes aux pages 40-43. Comme nous le remarquons plus haut, les éléments de théorie de la mesure exposés ici doivent être lus en parallèle avec Carnap (1960).

2) si $A \in CollA$ alors $A^c \in CollA$, (clôture par complémentation : si un événement A appartient à la collection, alors son complément aussi),

3) si $A, B \in CollA$ alors $A \cup B \in CollA$, (si deux événements appartiennent à la collection, alors leur réunion est aussi un élément de la collection),

Enfin, deux conditions pour une partition finie de S : A_i qui est un élément de $CollA$, où $i = 1, \dots, n$, on a :

4) $\bigcup_{i=1}^n A_i \in CollA$ (clôture sous réunions finies : toutes les réunions de A_i sont dans $CollA$).

5) $\bigcap_{i=1}^n A_i \in CollA$ (clôture sous intersections finies : toutes les intersections de A_i sont dans $CollA$).

Par contre, pour une partition infinie dénombrable de S A_i qui est un élément de $CollA$, où $i = 1, 2, 3 \dots$ etc. Nous parlons alors d'une σ -algèbre avec deux conditions supplémentaires :

6) $\bigcup_{i=1}^{\infty} A_i \in CollA$ (clôture sous réunions *dénombrables*).

7) $\bigcap_{i=1}^{\infty} A_i \in CollA$ (clôture sous intersections *dénombrables*).

Savage va utiliser une partition infinie¹ dénombrable, pour laquelle la mesure de probabilité doit être σ -additive, c'est-à-dire que :

8) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (ce qui signifie que, pour des parties de $CollA$ deux à deux disjointes, la mesure de la réunion des A_i est égale à la somme des mesures des parties de A_i).

¹ Pour une très bonne analyse des problèmes liés à l'utilisation des partitions finies voir, Kraft/Pratt/Seidenberg (1959) ; Fisher (1986), p 337 ; Kreps (1988), p 119 ; si S est fini, la mesure peut ne pas être unique, voir Gilboa (2009), p 111. C'est que montre un exemple de de Finetti, repris par Kraft/Pratt/Seidenberg (1959) avec un ensemble S divisé en 5 sous-ensembles, l'exemple est analysé par Kreps (1988), p 119.

Avec ces éléments mathématiques à l'esprit, revenons au texte de Savage, et au problème qu'il va analyser maintenant et qu'il exprime comme suit :

« If S carries a probability measure P and a qualitative probability \leq^* such that for every B, C , $P(B) \leq P(C)$ if and only if $B \leq^* C$ then P (strictly) agrees with \leq^* . If $B \leq^* C$ implies $P(B) \leq P(C)$ then P almost agrees with \leq^* »¹.

La question est donc de savoir par quels moyens mathématiques nous pouvons formaliser l'incertitude du sujet quant à la réalisation d'un événement quelconque, c'est-à-dire attribuer à ma croyance une mesure quantitative située dans l'intervalle $[0,1]$. C'est la fonction de probabilité elle-même qui va permettre la mise en rapport de P et \leq^* .

Nous allons ici étudier la démonstration de Kreps (1988)² que nous comparerons à celle de Savage. La démonstration de Kreps est moins générale mais beaucoup plus claire et concise que celle de Savage, et elle conduit absolument aux mêmes conclusions. Le but étant de découvrir : «...conditions under which there is a probability measure that agrees, either strictly or almost with a given qualitative probability »³. Afin de parvenir à ce but les deux auteurs vont utiliser des partitions sur un ensemble de référence donné. Il s'agit en fait d'utiliser ce que Fishburn (1986)⁴ nomme l'« axiome de partition 1 », et l'« axiome de partition 2 ».

L'idée générale de la méthode choisie par Savage est la suivante : préférant l'axiome 1, il utilise un événement B comme ensemble de référence dans lequel il s'agira de mettre en correspondance deux partitions *presque uniformes*⁵ B_i et C_j . Et c'est justement dans la mise en correspondance de ces deux partitions presque uniformes au moyen d'une transformation affine donnée que sera rendue possible le passage de \leq^* à P . En fait il faut imaginer B_i et C_j

¹ Savage (1954), p 34. L'auteur remarque aussi que si P s'accorde strictement avec \leq^* alors la connaissance de P implique celle de \leq^* . Mais des cas tératologiques peuvent se présenter où $P(B) = P(C)$ bien que $B <^* C$.

² Elle se situe aux pages 120-125.

³ Savage (1972), *ibid*.

⁴ p 341. Il y en fait trois axiomes et le troisième correspond à P6 chez Savage.

⁵ Ces partitions, qui sont des collections de sous-ensembles de B , c'est-à-dire des algèbres booléennes de B , ont pour caractéristiques principales que l'union de r éléments n'est pas plus probable que l'union de $r+1$ éléments. Les « tranches » de la partition peuvent donc être de tailles différentes. Rappelons que nous travaillons ici dans le cadre de partitions infinies dénombrables, dont les propriétés ont été définies plus haut.

comme des sortes de règles graduées – dont les graduations peuvent ne pas être égales – et où chaque élément C_j est la réunion de a ou $a + 1$ élément des B_i . L'existence de ces deux partitions permet à Savage de conclure qu'il existe *une et une seule* mesure P qui s'accorde presque avec \leq^* . C'est-à-dire que, pour tout $\rho, 0 \leq \rho \leq 1$ et tout $B \subseteq C$, et P unique tel que définit plus haut, il existe $C \subseteq B$ tel que $P(C) = \rho P(B)$. Autrement dit il existe un nombre réel ρ qui permet d'exprimer $P(C)$ en fonction de $P(B)$. C'est cette mise en relation de deux ensembles qui va permettre d'établir la mesure en tant que telle¹.

De son côté, Kreps utilise l'axiome de partition 2, c'est-à-dire une partition uniforme sur S en 2^n sous ensembles, c'est le choix que font aussi de Finetti (1937), Koopman (1940) et (1941). Il s'agit donc de partitionner S en 2^n événements égaux. La partition a la forme suivante :

$$A_1^n, A_2^n, \dots, A_{2^n}^n.$$

Elle a les propriétés suivantes³ :

- 1) $A_j^n \cap A_k^n = \emptyset$. Les éléments de la partition sont disjoints et indépendants.
- 2) $A_j^n \sim^* A_k^n$ pour tous les j et les k . Aucun événement n'est subjectivement plus probable qu'un autre.
- 3) $\bigcup_{i=1}^{2^n} A_i^n = S$. La partition est close pour les réunions dénombrables, autrement dit la réunion des $2^n A$ est équivalente à S .

Il convient à présent de se demander quelle probabilité p peut être assignée à chaque « tranche » A_j^n pour représenter \succ^* , et nous trouvons que pour tout j et tout n :

¹ Voir Granger (1988), p 256-259.

² 2^n représente l'ensemble de tous les sous-ensembles de S , autrement dit, l'ensemble de toutes les possibilités, n représente une ou plusieurs séries d'événements issus de l'ensemble des sous-ensembles de S . Autrement dit, nous pouvons nous imaginer ces éléments comme les 2^n séquences possibles de piles ou faces qui résulteraient de n lancers d'une pièce non biaisée. L'exemple de la pièce est de Savage (1954) p 33, qui l'élabore un peu plus, il est repris par Kreps (1988) dans son exposition p 120.

³ Nous allons utiliser le même symbole que Kreps : \succ^* qui a pour fonction de représenter une préférence stricte en termes de probabilité qualitative : $A \succ^* B$ si le sujet pense que A est strictement plus probable que B . Le fait que Kreps n'utilise pas le même symbole que Savage ne change rien à la portée de la démonstration.

$$p(A_j^n) = \frac{1}{2^n}$$

Si nous prenons maintenant un événement B , nous pouvons utiliser notre partition comme une règle graduée (cette fois de façon régulière, pas comme chez Savage) pour découvrir la probabilité de B . Nous procédons ainsi : pour un événement B et pour tout n de A_j^n il y a un k plus petit que n tel que :

$$\bigcup_{j=1}^k A_j^n \succ^* B$$

Comme k est directement relié à n , est fonction de n , on peut l'exprimer ainsi : $k(n)$. Dans ce cas, la probabilité de B , doit pouvoir être située dans l'intervalle :

$$\frac{(k(n)-1)}{2^n} \leq P(B) \leq \frac{k(n)}{2^n}$$

Nous arrivons ainsi à « coincer » la probabilité de B de façon de plus en plus fine selon que n tend vers l'infini. Autrement dit plus la partition est fine – plus le nombre de « tranches » est élevé – et plus nous nous rapprochons de $P(B)$, soit :

$$P(B) = \lim_{n \rightarrow \infty} \frac{k(n)}{2^n}$$

C'est exactement le résultat auquel parvient Savage, page 36, mais avec une hypothèse plus forte (axiome de partition 1) et une démonstration plus longue. Le rôle dévolu par Savage aux deux partitions B_i et C_j est ici dévolu à n et à 2^n qui constituent les unités de mesure de la partition. Dans l'approche de Kreps, comme dans celle de Savage, plus la partition est fine, et plus nous pouvons avoir une mesure précise. Les A_j^n pourraient être pensées comme les graduations d'une règle que l'on pourrait affiner autant que l'on veut. D'autre part, cette procédure garantit qu'il existe une probabilité unique pour représenter \geq^* .

Cependant, comme nous l'avons vu plus haut, il peut exister un cas tératologique où : $P(B)=P(C)$ bien que $B <^* C$, ce qui signifie que dans certains cas¹ $<^*$ ne peut être représenté par P . Il convient donc d'introduire deux clauses supplémentaires afin de s'assurer que ce cas ne puisse se présenter, ces deux clauses permettent de déduire un accord strict entre P et $<^*$.

1) La relation $<^*$ doit être « fine », c'est-à-dire que lorsque $\emptyset <^* B$ l'ensemble universel S peut être partitionné en événements C_1, C_2, \dots, C_n , tels que pour tous les membres de la partition $C_j \leq^* B$.

2) La relation \leq^* doit être « tight », c'est-à-dire que si $B^* \geq C$ il existe un D tel que $B^* \geq C \cup D^* \geq C$.

De ces deux clauses, mathématiquement développée et élaborées, Savage peut déduire les propriétés suivantes concernant la relation de probabilité qualitative : « If \leq^* is both fine and tight, the only probability measure that almost agrees with \leq^* strictly agrees with it, and there exist partitions of S into arbitrary many equivalent events »².

Voilà qui clôt l'analyse consacrée par Savage à la question du passage du qualitatif au quantitatif : nous disposons maintenant d'une mesure de probabilité P unique pour décrire la relation qualitative et subjective \leq^* . Nous disposons donc maintenant d'un outil pour *mesurer* de façon aussi précise que possible l'intensité des croyances du sujet sur une échelle qui peut être graduée de façon aussi fine que l'on veut, échelle située dans l'intervalle fermé $[0,1]$. Par ailleurs ces considérations permettent aussi à Savage, via le concept de partition, d'introduire son sixième axiome.

Axiome 6

« If $g < h$, and f is any consequence ; then there exists a partition of S such that, if g or h is so modified on any one element of the partition as to take the value f at every s there, other values

¹ Voir, Savage (1954), p 36-37. En fait il y a deux cas où la probabilité qualitative ne peut être représentée par une mesure quantitative, ils sont exposés de façon brillante par Kreps (1988), auquel nous renvoyons, p 122-123 ; voir aussi Fishburn (1986), p 341.

² Savage (1954), p 38.

being undisturbed ; then the modified g remains less than h , or g remains less than the modified h , as the case may require »¹.

Il s'agit d'un axiome de continuité, qui comme chez Ramsey, et N/M, repose sur une propriété archimédienne², pour plus de clarté, nous allons le formaliser comme suit : pour tout acte g et $h \in F$, tels que $h > g$, et pour toute conséquence $f \in F$, il existe une partition *finie* de S telle que pour tout A sur la partition :

$$[h'(s)=f \text{ pour } s \in A, h'(s)=h(s) \text{ pour } s \in A^c] \text{ implique } h' > g$$

$$[g'(s)=f \text{ pour } s \in A, g'(s)=g(s) \text{ pour } s \in A^c] \text{ implique } h > g'$$

Notons tout d'abord que la partition sur S doit être *finie*, ce qui n'était pas le cas dans le développement sur les probabilités subjectives. Comme le fait remarquer Gilboa (2009)³, l'axiome 6 et l'axiome 7 sont des clauses techniques qui n'ont pas une très grande portée normative⁴. Elles sont donc certes intéressantes à étudier du point de vue des mathématiques de la décision, mais moins intéressantes du point de vue qui est le nôtre et qui concerne surtout les prétentions normatives de la TUA.

Toutefois quelques remarques s'imposent. Tout d'abord, comme nous l'avons vu chez N/M, plus que chez Ramsey peut-être, l'axiome de continuité sert à garantir qu'aucune conséquence ne puisse être indéfiniment désirable ou indésirable en elle-même. Il y a donc ici un aspect nettement limitatif⁵. La propriété archimédienne, qui chez N/M était reliée à la

¹ Avant d'introduire P6, Savage avait eu l'idée d'introduire un axiome P6', moins général qui stipulait que si $B < C$ il existe une partition de S dans laquelle la réunion de chaque élément de S avec B est moins probable que la réunion de chaque élément de S avec C . Il s'agit d'une forme d'axiome de monotonie par rapport aux partitions. Pour une bonne explication voir Fishburn (1970), p 194. Cet axiome, bien que suffisant dans le cadre de la théorie des probabilités que Savage vient de développer, ne l'est plus lorsque nous faisons intervenir le concept d'utilité sur les conséquences. Il faut donc un axiome plus général, ce sera P6.

² Pour cette propriété, voir Ramsey (1926), p 171. En fait il s'agit ici de trouver un h' ou un g' assez proche de h et g de telle sorte que les préférences initiales ne soit pas modifiées, si on remplace par exemple, h par h' .

³ Page 103, voir aussi Shafer (1986), p 468.

⁴ Avis qu'il faut toutefois modérer quand on sait la portée descriptive de l'axiome de continuité chez N/M. La façon dont Savage présente sa propre clause de continuité masque un peu son aspect descriptif, mais ce qui était vrai chez N/M, et chez Ramsey, l'est aussi chez Savage.

⁵ Voir, Fishburn (1970), p 194 ; Kreps (1988), p 131-132, encore une fois soulignons la pertinence pédagogique de l'exposition de cet auteur ; enfin Gilboa (2009), p 103-104 donne une excellente analyse de P6 en le mettant en rapport avec les éléments de théorie de la mesure que nous venons d'étudier.

distribution de probabilité, est ici liée à la taille de la partition¹. Ce qui signifie que pour une partition suffisamment fine, une conséquence f (favorable ou non) peut être intégrée parmi les conséquences possibles d'un acte h sans que les préférences par exemple entre h et g ne soient modifiées. Autrement dit, pour chaque partie d'une partition de S (aussi fine que l'on veut), si nous modifions h en le rendant identique à f (h') sur la partie en question, cette partie sera tellement petite – en termes de probabilités – que la préférence originelle sur les actes ne sera pas modifiée. Il en ira de même pour les actes g et g' ². Nous remarquons aussi que bien que les façons de construire et d'introduire la clause de continuité ne soient pas identiques chez Ramsey, N/M et Savage, l'intuition qui sous-tend ces axiomes est proche dans les trois cas.

L'axiome 7 et le problème de l'utilité

Le concept d'utilité est l'objet du cinquième chapitre, la notion d'utilité proprement dite est introduite dans la troisième partie qui porte sur l'ordonnement des préférences entre des paris³ :

« A *utility* is a function U associating real numbers with consequences in such a way that if $\mathbf{f} = \sum \rho_i f_i$, and $\mathbf{g} = \sum \sigma_j g_j$; then $\mathbf{f} \leq \mathbf{g}$, if and only if $\sum \rho_i U(f_i) \leq \sum \sigma_j U(g_j)$. Writing $U[\mathbf{f}]$ for $\sum \rho_i U(f_i)$, the condition takes the form $U[\mathbf{f}] \leq U[\mathbf{g}]$. Similarly it is convenient to understand, that, for an act \mathbf{f} , $U[\mathbf{f}] = E(U[\mathbf{f}])$ »⁴.

L'utilité est donc d'abord définie pour les paris⁵, c'est-à-dire des actes aux conséquences finies⁶. Dans ce cadre-là ce sont les probabilités attachées à chaque conséquence

¹ Remarquons que du point de vue de la structure, les clauses de continuité de N/M sont différentes de celles de Savage en ce sens qu'elles font intervenir trois alternatives, alors que chez Savage il s'agit de faire jouer entre eux deux actes, leurs modifications sur une partition, et une conséquence.

² P6 est aussi par là même introduit pour assurer l'existence d'un ordre à valeurs réelles préservant les probabilités.

³ Voir, Savage (1954), p 73. Comme la démarche de notre auteur est très proche de celle de N/M, nous n'insisterons pas beaucoup sur cette partie. Remarquons toutefois que dans Savage, la notion de loterie, à partir de laquelle était calculée l'utilité chez N/M, est remplacée par le concept de pari qui permet d'introduire les probabilités subjectives. Sur le concept de pari, voir Granger (1967), p 69-72, et Hacking (2004), p 127 et p 178.

⁴ Savage (1954), p 73. Le symbole \mathbf{f} représente un pari aux conséquences finies, ρ et σ sont les probabilités, f et g les conséquences.

⁵ La définition mathématique du pari se trouve p 71.

⁶ Les conséquences des actes doivent être finies pour que l'utilité le soit aussi.

singulière qui vont déterminer les préférences. Nous sommes donc très proches de la solution N/M. Il s'agit pour Savage de déduire une équivalence entre l'utilité espérée $U(f)$ et la mesure de probabilité \leq . C'est ce que formalise le théorème 1 : pour qu'il existe une fonction d'utilité de F (ensemble des actes¹) dans \mathbb{R} il faut que $f \leq g$ soit équivalent à $U[f] \leq U[g]$ ².

La valeur d'un pari est calculée grâce à la formule $\sum \rho_i f_i$ qui donne l'utilité attendue de l'acte aux conséquences finies f . La formule qui conclut notre citation signifie que l'utilité d'un acte est équivalente à son utilité attendue. À partir de là et des axiomes P1-P6, il est possible de déduire une fonction d'utilité de F (ensembles des conséquences) dans \mathbb{R} qui peut être caractérisée ainsi :

$$(f \leq g) \equiv E[U(f(s))P^*] \leq E[U(g(s))P^*] \quad ^3.$$

Comme nous le notions plus haut, nous sommes ici dans le cadre d'actes aux conséquences finies, c'est-à-dire des paris. D'autre part, nous remarquons que les six premiers axiomes *suffisent* pour déduire une fonction d'utilité. Une question dès lors se pose : pourquoi Savage a-t-il introduit un septième axiome, dont l'énoncé est le suivant :

« If $f \leq (\geq) g(s)$ given B for every $s \in B$, then $f \leq (\geq) g$ given B »⁴.

Cet axiome⁵ qui est à mi-chemin entre la monotonicité⁶ et l'indépendance sert à Savage à appliquer la notion d'utilité telle que définie plus haut à des classes d'actes plus larges que les paris. L'axiome formalise l'intuition suivante, qui peut sembler triviale : si un acte f est faiblement préféré à une conséquence $g(s)$ étant donné B , pour tous les états du monde inclus dans B , alors l'acte f est faiblement préféré à l'acte g sous lequel se range la conséquence $g(s)$. Quand Savage parle d'actes plus généraux, il exclut par contre les actes dont l'utilité des

¹ C'est justement P7 qui va permettre de passer des paris à des actes plus généraux.

² U doit représenter complètement, de par sa valeur, la relation de préférence sur les actes. Voir Picavet (1996), p 215.

³ Nous pouvons parler ici d'utilité *subjective* des perspectives. P^* représente d'ailleurs la mesure de probabilité subjective attribuée à la perspective.

⁴ Savage (1954), p 77.

⁵ Dont l'auteur dit qu'il est dans « l'esprit du PCS », probablement parce qu'il formalise lui aussi une forme d'indépendance des préférences.

⁶ Le qualificatif est de Karni (2005), p 7 ; Shafer (1986), p 468 parle lui de « dominance ».

conséquences est infinie¹, tels ceux qu'on trouve par exemple dans le pari de Pascal². La fonction d'utilité sur F doit être limitée. En fait l'axiome 7 sert à démontrer que l'équation présentée plus haut vaut pour *tous* les actes dont l'utilité est finie. Il s'agit donc bien d'une clause à caractère technique³ qui n'a pas de portée normative, mais seulement une portée limitative concernant l'axiomatique et la fonction d'utilité.

Nous concluons cette analyse de la théorie de Savage par quelques remarques d'ordre et de portées différentes. Tout d'abord nous tenons à insister sur la remarquable cohérence mathématique de l'entreprise, et sur la remarquable unité de l'axiomatique. Bien que les axiomes puissent être divisés en deux groupes (1-4, 5-6), normatifs et techniques, c'est l'unité de la structure et du mouvement de déplacement de la relation de préférences des actes vers les événements qui fait l'originalité de la théorie de Savage. Par ailleurs il faut bien garder à l'esprit, et nous pouvons ici reprendre les remarques qu'Oskar Morgenstern faisait à propos des critiques de la théorie de l'utilité N/M⁴, que c'est rarement sur la consistance interne de l'axiomatique en tant que telle, que les critiques porteront, mais sur des axiomes en quelque sorte « décontextualisés ». Ici l'axiome d'indépendance⁵, là l'ordonnancement et la transitivité⁶, qui sont des clauses centrales. Loin de nous l'idée de vouloir minimiser la portée de ces critiques, qui vont justement nous servir à construire nos propres hypothèses, mais il faut juste remarquer qu'elles restent d'une certaine façon extérieures et n'atteignent pas la consistance interne des théories de N/M et Savage.

Comme nous le notions plus haut, c'est la *portée normative* du PCS et de l'axiome 2 qui seront soumises à l'examen le plus radical, et ce pour la raison que ces clauses représentent le « coeur » de la TUA et sont au fondement de l'idée de maximisation. Savage avait pour but de construire une théorie normative, il fait part de cette idée à plusieurs

¹ En fait, les axiomes P1-P6 excluent déjà par eux-mêmes ce type d'actes.

² La littérature sur ce passage de Pascal est plus qu'abondante en philosophie, voir Nadal (1980), pour une approche probabiliste du problème ; Gouhier (1984) ; Carraud (1992) ; cependant l'interprétation du Pari du point de vue de la théorie de la décision a été le fait de Granger (1967), et bien entendu Hacking (1972), (1975), p 63-72, et (2004), p 127.

³ Comme le remarque Gilboa (2009), p 104.

⁴ Voir Morgenstern (1979), in Allais/Hagen (1979), p 175-183.

⁵ C'est le cas d'Allais (1951), et de Kahneman/Tversky (1979).

⁶ C'est le cas d'Edwards (1954), et de Lichtenstein/Slovic (1968) et (1971).

reprises¹, une théorie qui puisse servir aux sujets en contexte décisionnel à amender leurs choix au cas où celui-ci soit justement en contradiction avec les normes de la rationalité telles que l’auteur les propose. Rappelons que le calcul des probabilités et le respect par le sujet de ses axiomes ainsi que les clauses P1-P7 représentent pour Savage l’essence même de la rationalité normative².

Un dernier point doit attirer notre attention, et il est relié à l’aspect normatif de la théorie que nous avons discuté plus haut. En 1952, lors d’un *Colloque international sur le risque* organisé à Paris, l’économiste français Maurice Allais, proposa aux participants, incluant Savage, une matrice de choix ayant la forme suivante :

« (1) *Préférez-vous la situation A à la situation B ?*

Situation A : Certitude de recevoir 100 millions.

Situation B : 10 chances sur 100 de gagner 500 millions.

89 chances sur 100 de gagner 100 millions.

1 chance sur 100 de ne rien gagner.

(2) *Préférez-vous la situation C à la situation D ?*

Situation C : 11 chances sur 100 de gagner 100 millions.

89 chances sur 100 de ne rien gagner.

Situation D : 10 chances sur 100 de gagner 500 millions.

90 chances sur 100 de ne rien gagner. »³

Il s’agit là de ce qui deviendra dans un avenir bref le fameux « paradoxe d’Allais » que nous allons analyser sous peu. Or quelle fut la réponse de Savage à cette situation ? Il émit les préférences suivantes : $A > B$ et $D > C$ ⁴. Or ces préférences sont des violations claires des axiomes d’indépendances de N/M, du PCS et de P2 de Savage⁵, et ce dans deux dimensions :

¹ Notamment à la page 97 de son ouvrage.

² Ramsey attribue aussi aux probabilités un aspect normatif. Savage consacre une partie de son ouvrage à ce problème aux pages 56-68, la thèse de la normativité des probabilités par Savage, et le « code de conduite » qu’elles constituent sont analysées par Saint Sernin (1973), p 44-48.

³ Allais (1953) p 527.

⁴ Savage (1954) discute de façon approfondie cette question aux pages 102-104, et c’est à ce propos qu’il va avoir aussi une autre intuition fondamentale : celle de l’effet de cadrage (framing effect), effet qui signifie que la façon dont les problèmes sont présentés influe sur les choix des sujets, ce qui est une violation du principe d’invariance qui lui stipule que les préférences doivent être indépendantes du contexte. Voir Kahneman/Tversky (2002), p 209-240.

⁵ Préférences qui devraient être : si $A > B$ alors $C > D$. C’est à partir de ce schéma que Kahneman et Tversky vont établir leur Prospect Theory.

la règle de maximisation de l'utilité attendue n'est pas respectée, et le principe d'indépendance est violé. Bien que sentant une « grande attirance intuitive pour ces préférences », Savage en est venu à les amender en fonction de son axiomatique et du PCS, en construisant le paradoxe¹ d'une façon qui laissait moins de place à la possibilité de se tromper. C'est donc ainsi que Savage concevait le rôle de son axiomatique, et il a lui-même illustré, de par sa propre expérience, comment et pourquoi les sujets devaient s'y confirmer, ce qui demande tout de même un peu de réflexion et d'élaboration².

Enfin, l'idée par laquelle Savage commence sa théorie et qui est la suivante : le monde pertinent pour la décision doit être découpé en fonction de la situation décisionnelle, autrement dit, le monde pertinent doit être un « small world », anticipe d'une façon assez nette nos vues sur la construction de l'*environnement décisionnel* ou *visée* par le système de référence décisionnel. Mais une différence importante demeure : Savage se situe dans le registre du formalisme, formalisme auquel nous serons obligés de tourner le dos, au moins partiellement, pour rendre plus réalistes nos hypothèses.

¹ Certains auteurs pensent que les violations de la TUA produites par le paradoxe sont liées à la façon dont il est construit, voir Birnbaum/Schmidt (2010) ; Amihud (1979), in Allais/Hagen, p 152-153.

² C'est un gros problème concernant l'applicabilité des axiomatiques aux situations réelles.

Chapitre II

Du paradoxe d'Allais à la Théorie de la Perspective :

L'introduction de la dimension subjective dans le choix rationnel.

Introduction

La critique de la TUA opérée par les articles que nous allons étudier dans ce chapitre et en particulier la Théorie de la Perspective de Kahneman et Tversky (1979), vont nous permettre de discuter plusieurs éléments de la théorie classique de la décision, et notamment les questions liées à la *normativité* de cette théorie. Mais ce n'est pas la raison principale qui nous a poussés à étudier les travaux d'Allais, Edwards, Lichtenstein/Slovic (L/T) et enfin Kahneman et Tversky (K/T). En effet, les paradoxes que nous allons examiner, et des phénomènes comme le renversement des préférences¹, en plus de mettre au jour une forme d'incohérence² du point de vue de la rationalité, démontrent de façon très claire que dans l'élaboration des choix rationnels intervient une strate subjective difficilement formalisable, puisqu'elle échappe aux axiomatiques³. Néanmoins cette strate subjective est difficilement éliminable si la théorie de la décision doit être réaliste⁴.

Dans la cadre de nos travaux et de nos hypothèses, cette strate est essentielle puisque c'est son fonctionnement que nous voulons mettre au jour dans le travail du SRD. Par ailleurs, ce ne sera que par et dans la critique que, progressivement, les éléments subjectifs vont apparaître, notamment par le truchement de phénomènes tels que la pondération des probabilités qui va être analysée par Allais et Edwards, puis Kahneman et Tversky. Le phénomène de pondération des probabilités, qui mènera à l'élaboration du coefficient π dans la TP, ne peut être expliqué que par l'effet d'une *représentation subjective* des probabilités. *Représentation* et *subjectivité* sont pour nous les deux concepts clés au moyen desquels nous allons reconstruire le concept de volonté, puisque nous faisons l'hypothèse que *la volonté c'est la décision dans l'élément de la représentation*.

La possibilité, pour les articles que nous allons lire, de faire émerger à partir de la critique de la TUA les éléments nécessaires à la construction de notre hypothèse justifie ainsi donc, de façon presque négative, les plans des chapitres I et II, car c'est seulement par et dans

¹ Phénomène qui sera mis en avant par Slovic et Lichtenstein (1968) et (1971).

² La question de savoir si les sujets sont réellement incohérents en ne choisissant pas selon les axiomes de la TUA reste aussi largement ouverte, comme en témoigne les analyses de Simon (1955) et les analyses regroupées dans Gigerenzer (Ed.) (2011) sur la rationalité limitée. Il semblerait que le fait de ne pas obéir aux axiomes puisse être versé au registre d'une rationalité limitée et adaptative, et non pas nécessairement à une défaillance. Nous sommes assez en accord avec les vues de Simon et Gigerenzer.

³ Même si ces axiomatiques iront en se raffinant.

⁴ C'est un point sur lequel insistera Allais (1953).

la critique que la subjectivité et la représentation vont faire progressivement leur entrée dans le domaine de la théorie du choix rationnel, même si c'est au prix du paradoxe.

Un aspect central de tout choix rationnel, que la critique va mettre en avant, est que ce qui importe aux yeux du sujet – contrairement à ce que postulait la TUA –, ce n'est pas la seule considération de l'avoir futur, mais le poids et l'importance de l'avoir présent¹. Cela est à relier au fait que tout choix rationnel est connecté à ce que nous présenterons dans notre troisième chapitre comme le Système de Référence Général (SRG), à savoir les éléments cognitifs qui déterminent et définissent le sujet du point de vue représentationnel et cognitif au moment de son choix. Il ne s'agit pas ici de dire que le sujet ne considère plus les conséquences possibles de ses choix, ce qui serait absurde, mais de dire qu'il ne les considère qu'à la lueur de son avoir présent. Nous parlons ici d'avoir et non d'être, car nous restons encore dans le domaine de l'économie.

C'est la théorie de la Perspective de Kahneman et Tversky qui a permis de formaliser le rôle et l'impact que l'avoir présent peut avoir sur les décisions par le biais de la fonction de valeur v . C'est un concept central pour nous, car c'est à partir de lui que nous construirons le concept d'*état présent* du Système de Référence Décisionnel (SRD). D'autre part, mesurer comment les perspectives sont évaluées à partir de l'avoir présent des sujets semble plus réaliste que de partir du principe que seule l'utilité attendue des résultats intervient.

La Théorie de la Perspective de Kahneman et Tversky représente une synthèse qui récapitule et intègre dans son formalisme la plupart des critiques² qui avaient été adressées à la théorie de l'utilité. Elle a été élaborée aux termes d'une évolution qui commence dans les travaux de Markovitz (1952), et qui passe notamment par les articles que nous allons étudier, elle coïncide aussi avec deux éléments importants : la naissance d'une psychologie cognitive de type scientifique³, et l'introduction du bayesianisme dans le domaine de la psychologie de la décision par Edwards, Lindman et Savage⁴.

Quel rôle joue la Théorie de la Perspective dans le cadre de notre travail ? C'est grâce à elle qu'ont été introduits et formalisés les éléments centraux que sont pour nous des concepts

¹ C'est ce que va mettre en avant la Théorie de la Perspective.

² En effet, il y a des phénomènes que la TP ne parvient pas à formaliser comme le phénomène de la dominance stochastique qui est présent dans le paradoxe d'Ellsberg par exemple.

³ Avec une élaboration des méthodes d'investigations empiriques et des méthodes statistiques d'analyse des résultats.

⁴ L'article de référence dans le domaine étant : « Bayesian Statistical Inference for Psychological Research » (1963).

comme le coefficient de pondération, la fonction de valeur, les phases d'édition et d'évaluation, et enfin les concepts d'aversion au risque et de statu quo. Ces éléments sont centraux, car nous allons tous les exploiter philosophiquement parfois en généralisant leur portée, et nous pensons qu'ils sont le noyau de toute décision humaine. Seulement, ce que la Théorie de la Perspective nous offre c'est certes le cœur de toute décision, mais c'est un cœur formel qu'il va falloir incarner.

Le fait que la Théorie de la Perspective ait pour vocation de résorber les critiques adressées à la TUA en restant aussi proche que possible de cette dernière en fait pour notre propos une théorie certes centrale mais insuffisante et incomplète pour plusieurs raisons. Par exemple elle ne prend que très peu en compte l'aspect essentiellement dynamique de toute décision, elle ne permet pas non plus de penser un processus décisionnel aussi important que l'établissement du but et de la visée. Aussi dans notre optique, qui est de proposer une théorie de la volonté à partir des éléments acquis dans la Théorie de la Perspective, il conviendra d'opérer une réorganisation et un réajustement de l'ensemble de ces éléments dans un cadre philosophique élargi : le Système de Référence Décisionnel.

A – Les premières critiques de la TUA.

1) La valeur psychologique des perspectives et le problème de la pondération.

Les théories de Ramsey, N/M et Savage sont unanimes sur le point suivant : tout recourt à l'introspection dans le cadre de l'analyse des processus décisionnels doit être éliminé¹. Mais plus largement, c'est toute la dimension psychologique du décideur qui doit être ignorée². La TUA déconnecte à la fois le décidable du volontaire³ et le psychologique du rationnel : le rationnel – c'est-à-dire dire le respect des axiomes – exclu de son expression l'ensemble des déterminations et des éléments psychologiques qui interviennent dans les choix⁴.

C'est précisément autour de ce problème que va s'articuler la critique de la TUA menée par l'économiste français Maurice Allais dans son article : « Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine »⁵. Les éléments que l'économiste va mettre en place dans son texte sont centraux en ce qui concerne la « psychologie du risque »⁶ ainsi que pour la suite de notre travail, et ce dans deux

¹ Voir Edwards (1954), p 385.

² La TUA est une modélisation strictement mathématique. Cependant, Ramsey (1926), et Savage (1954), p 103, font néanmoins de légères incursions dans le domaine de la psychologie.

³ Elle partage d'ailleurs cette particularité avec les théories économiques et psychologiques que nous allons étudier. Et cette carence nous pose un problème, car comment penser le décisionnel en faisant abstraction du volontaire ? Il conviendra dans la suite de notre travail, au troisième chapitre, de réinjecter le concept de volonté, repensé et reconstruit, dans le domaine du décidable.

⁴ Une des prémisses de la TUA, que met bien en avant la théorie de N/M, c'est qu'il faut décontextualiser le choix rationnel le plus possible du point de vue psychologique, afin de pouvoir construire des situations décisionnelles qui soient complètement épurées. C'est d'ailleurs en partie à cause de cela que N/M éliminent – avec leur axiome 3 : C : b – des éléments comme le plaisir pris au jeu. Or nous savons justement, de par l'introspection que chacun peut faire et grâce aux avancées de la psychologie cognitive, que dans la plupart de nos décisions interviennent de multiples facteurs de types psychologiques mais aussi de types environnementaux. La littérature sur ce point est particulièrement abondante, mais par exemple, la redécouverte de l'effet de cadrage (« framing effect ») par Kahneman et Tversky (1979), (1986), semble indiquer que la façon dont le sujet perçoit les options est fondamentale. Pour un point de vue général sur ce problème, voir Baron (1988), p 229 - 524 ; et Koehler/Harvey (2004), p 1-133 et 339-504.

⁵ *Econometrica*, (1953), Vol. 21, No. 4. Cet article est en fait la reprise un peu abrégée d'un texte beaucoup plus long écrit en 1952 : « The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the american school » réédité en 1979 à l'occasion de la parution du livre important de Allais/Hagen : *Expected utility and the Allais paradox*. Dans l'article de 1952, la théorie de l'utilité d'Allais est élaborée plus longuement que dans l'article de 1953 qui est essentiellement axé sur la critique de la TUA. Nous citerons ces deux articles : Allais (1952), Allais (1953). Nous utiliserons certains textes publiés dans l'ouvrage mentionné plus haut, notamment ceux de Y. Amihud, et d'O. Morgenstern, qui sont deux critiques intéressantes des points de vue d'Allais. Par ailleurs la bibliographie à laquelle l'article de 1953 et le « paradoxe » ont donné lieu est bien trop importante pour en faire une revue ici. Nous introduirons donc les références pertinentes au fur et à mesure.

⁶ Les guillemets sont de rigueur, car il y a parfois des flottements concernant ce qu'Allais entend par psychologie et par rationalité. Il sera d'ailleurs critiqué à ce propos par Edwards dans son article de 1954, p 394, que nous étudierons par la suite.

dimensions. Tout d'abord, l'article d'Allais anticipe sur bien des points la Prospect Theory de Kahneman et Tversky¹ que nous allons étudier sous peu. Ensuite, la réflexion d'Allais sur la notion de pondération² est importante, car la thèse que nous soutenons dans ce travail a pour vocation de montrer que le concept de pondération, qui doit être généralisé de façon adéquate, par exemple au temps, est un élément essentiel de toute prise de décision, voire même l'élément essentiel ; mais aussi qu'une forme de pondération est à l'œuvre dans l'opération de catégorisation du SRD³. L'idée, qui sera développée dans le troisième chapitre – après avoir réintroduit le volontaire dans le décidable – est que s'il doit y avoir des processus à l'œuvre dans l'exercice du SRD, ils sont du type pondérateur et évaluatif⁴.

Par ailleurs il faut remarquer que les quatre facteurs centraux dans toute prise de décision en contexte de risque que va présenter Allais dans la première partie de son article émergent à partir de la critique de la TUA concernant la négligence des éléments psychologiques intervenant dans tout choix rationnel. Le but d'Allais est de proposer une théorie qui rende compte – au moins partiellement – de l'épaisseur de la subjectivité, mais nous aurons l'occasion de nous apercevoir, que cette théorie demeure incomplète du point de vue philosophique. Cette critique de la TUA par le moyen de la contestation empirique permet à Allais d'inaugurer⁵ le domaine d'études qui nous conduira à la Prospect Theory. C'est la critique de la TUA qui fera le lien et la continuité entre les travaux d'Allais, d'Edwards, Lichtenstein/Slovic⁶ et enfin Kahneman/Tversky.

Dans son article, Allais dénombre quatre facteurs fondamentaux qui doivent être pris en considération dans l'analyse des décisions en contexte de risque, facteurs qui selon l'auteur

¹ Dans la première partie de leur article ces auteurs vont d'ailleurs se référer de façon assez massive à Allais, voir en particulier, Kahneman/Tversky (1979), p 20-22.

² Cette notion est introduite d'emblée, p 504. Nous montrerons dans la suite de nos analyses que la pondération doit être considérée comme une catégorie de la représentation, catégorie dont la vocation est de transformer, de distordre, un ou plusieurs éléments du contexte décisionnel.

³ Rappelons ici que dans notre optique le SRD est équivalent à l'acte de la volonté.

⁴ Nous avons l'intention de reconstruire le concept de volonté en étendant la portée des phases d'édition et d'évaluation de la TP, ainsi que grâce aux concepts de fonction de valeur (« value function ») et de coefficient de pondération (« weighting function ») élaborés par Kahneman et Tversky justement à partir des travaux d'Allais. Et ces deux éléments ne pourront émerger qu'après une critique radicale de la TUA.

⁵ En fait c'est Mosteller/Nogee (1951) qui ont commencé à tester de façon empirique la TUA, voir Picavet (1996), p 229-266.

⁶ Avec cette réserve toutefois que les travaux d'Edwards nous font entrer dans le domaine de la psychologie bayésienne et de la critique de l'économie, voir Edwards (1954) et (1961). En ce qui concerne Lichtenstein et Slovic (1968) et (1971), le but originel de leurs travaux n'était pas une critique de la TUA, la critique fut un effet collatéral. Il n'en demeure pas moins que les acquis des expériences de ces auteurs sont centraux pour l'ordonnement des préférences et la transitivité, voir Lichtenstein/Slovic (2006).

sont ignorés par la TUA¹. D'autre part, la critique de la TUA va se constituer dans deux plans : la rationalité abstraite et la rationalité concrète. Les quatre facteurs sont tous reliés au concept central de « valeur psychologique » qui doit lui-même être associé à chaque gain monétaire² et qui doit être défini comme un seuil psychologique. D'un point de vue formel, ce concept a son origine dans la loi de Fechner-Weber³. Il s'agira donc ici de relier l'utilité cardinale $\bar{s}(x)$ au concept de *degré* de satisfaction⁴ et non plus à l'idée que les sujets doivent maximiser une certaine fonction – qu'Allais appelle un indicateur – $B(x)$. Mais plus profondément il va s'agir de montrer que dans les quelques cas particuliers où l'indicateur $B(x)$ existe, il ne peut être différent de $\bar{s}(x)$ à moins de négliger un facteur essentiel de toute prise de décision : la dispersion des valeurs psychologiques⁵. Allais soumet à une critique assez radicale la formulation canonique de la TUA, ou « formulation de Bernoulli », qui formalise l'essence même de la TUA, c'est-à-dire l'idée de maximisation. Formule qui est la suivante :

$$B(V) = \sum p_i B(g_i)$$

Or, nous avons vu que cette hypothèse est commune aux axiomatiques de N/M et Savage⁶. Il s'agira de montrer qu'un sujet rationnel⁷ ne construit pas ses choix en obéissant à la formule de Bernoulli. Pour mener à bien cette critique seront discutés le PCS de Savage, l'axiome de substituabilité de Samuelson⁸ et le comportement de sujets prudents face à des paris mettant en jeu de petites sommes. L'idée générale qui commande la critique de

¹ Il y a aussi toute une série de facteurs secondaires dont Allais dit qu'ils ne rentrent pas en ligne de compte lorsque l'on parle de « psychologie pure du risque », voir Allais (1953), p 506. Ces facteurs sont par contre analysés de façon exhaustive dans l'article de 1952, p 57-60. Par exemple il s'agit de l'auto justification du sujet face à une décision particulière, de la relation aux autres, du plaisir de gagner, etc. Ces facteurs qu'Allais qualifie de secondaires, sont très intéressants d'un point de vue philosophique, et nous aurons l'occasion de montrer dans notre troisième chapitre que certains d'entre eux jouent en fait un rôle de premier plan.

² Puisque nous sommes toujours ici dans une perspective économique.

³ Cette loi décrit la sensation comme une fonction logarithmique de l'intensité de la stimulation. Elle permet à Fechner (1860) d'établir le concept de « seuil différentiel » qui va être utilisé ici par Allais pour formaliser l'intuition d'une quantité minimale perceptible de satisfaction. Comme l'indique une note de Marschak, in Allais/Hagen (1979), p 173, il y a une filiation très nette de Bernoulli (1737) à Fechner (1860) en ce qui concerne l'utilisation de la fonction logarithmique pour formaliser le concept de seuil. Pour une analyse de cette filiation on peut aussi se reporter à Stigler (1950), p 374-377.

⁴ Remarquons que ce concept qui n'apparaît pas dans la TUA a d'emblée une portée plus psychologique.

⁵ Voir, Allais (1953), p 505 et p 535, pour un point de vue plus complet sur ce point voir Allais (1952), p 74-75.

⁶ Auxquelles Allais rajoutent les théories de Friedman/Savage (1948), Marschak (1950), Samuelson (1952), théories qui sont dans la stricte continuité du paradigme N/M. La théorie de Samuelson doit être l'objet d'une attention particulière, car c'est à travers la critique de l'axiome de substituabilité – qui est une forme affaiblie de l'axiome d'indépendance – qu'introduit cet auteur qu'Allais atteindra par ricochet le cœur de la théorie de N/M.

⁷ Nous aurons l'occasion de nous rendre compte que les définitions de la rationalité que propose Allais laissent parfois à désirer.

⁸ Rappelons la forme de cet axiome : $(\forall u)(\forall v)(\forall w) \in U [(u \sim v) \supset [(u1/2) + (1/2w) \sim (v1/2) + (1/2w)]]$.

l'économiste est la suivante : « [l'utilité] peut être déterminée seulement par *l'observation introspective* de différences de niveaux de satisfaction égale en fonction d'un seuil psychologique sensible »¹.

Le premier élément² (« élément I ») intervenant dans les décisions en contexte de risque qu'Allais introduit est le suivant : « La déformation psychologique des valeurs monétaires et la courbure de la satisfaction absolue »³. Il s'agit ici d'un phénomène général de pondération que Bernoulli (1737) décrivait déjà comme le phénomène de diminution de l'utilité marginale⁴. Ce phénomène signifie qu'il n'y pas une adéquation nécessaire entre la valeur monétaire g et la valeur psychologique $\bar{s}(g)$, autrement dit $\bar{s}(g)$ n'est pas une fonction linéaire de g ⁵. D'autre part, il faut noter que l'utilité marginale n'a bien entendu pas la même valeur selon l'état de fortune du sujet : un pauvre n'attribuera pas la même utilité marginale à un dollar qu'un riche ou même qu'un homme dont l'état de fortune est dans la moyenne. Ce phénomène qui avait été pressenti par Bernoulli⁶ est au fondement du concept de fonction de valeur (« value function ») qui sera développé par Kahneman et Tversky dans la théorie de la perspective⁷. Nous utiliserons ce concept dans notre troisième chapitre pour construire la notion de système de référence décisionnel.

Nous avons donc affaire ici à un premier phénomène de pondération en fonction de l'avoir et de l'état de fortune présent du sujet, pondération qui va nous amener à nous interroger sur le rôle joué par la représentation dans les processus décisionnels. Représentation dont nous pouvons déjà pressentir le rôle central, rôle qui sera d'ailleurs confirmé dans deux domaines : en théorie de la perspective, la fonction de valeur et le coefficient de pondération peuvent être pensés comme des effets de la représentation⁸, par ailleurs le lien intime qui unit

¹ Allais (1953), p 505. Comme nous remarquons plus haut, en faisant intervenir l'introspection nous nous éloignons immédiatement de la TUA.

² Tous les éléments qui seront étudiés ont un rapport avec les concepts de pondération et de représentation.

³ Allais (1953), p 507 ; voir aussi Allais (1952), p 47.

⁴ La notion d'utilité marginale est analysée dans les articles classiques de Vickrey (1945), Stigler (1950), et surtout Kauder (1965) qui développe un point de vue historique très intéressant.

⁵ C'est ce que montre l'exemple que développe Allais dans la note 6 de la page 507. En fait le centre de gravité du choix rationnel est le concept de valeur psychologique.

⁶ Bernoulli (1737), p 26.

⁷ Voir, Kahneman/Tversky (1979), p 32-34.

⁸ Voir l'article séminal de 1979, p 32-38. Un des buts de la Prospect Theory sera justement de formaliser la façon dont la représentation subjective distord les probabilités et les valeurs. Granger (1967), a eu une forte intuition anticipatrice en ce qui

volonté et représentation est connu au moins depuis Platon, Aristote, Saint Augustin, en passant par Schopenhauer et Ricoeur¹. Nous faisons d'autre part l'hypothèse que c'est par le concept de représentation que nous parviendrons à lier de nouveau le décidable et le volontaire dans notre troisième chapitre.

Un des grands mérites d'Allais – que suivront ici Kahneman et Tversky – est d'avoir, du fait de la prise en compte de la psychologie du sujet, réintroduit le concept de représentation dans le domaine de la théorie du choix rationnel grâce au concept de pondération. Ce mérite est toutefois limité en ce sens que l'auteur n'a pas complètement explicité la portée philosophique de cette réintroduction. Par ailleurs il faut remarquer que les éléments liés à la représentation sont complètement ignorés par la TUA puisqu'ils ont affaire avec ce que le sujet « ressent » face à une option donnée. Et précisément, ce ressentir est hors du domaine de la TUA.

Le deuxième facteur² (« élément II ») qu'Allais va analyser est central pour tout le reste de notre travail, il s'agit de : « *La déformation subjective des probabilités objectives* »³. Il s'agit ici de conceptualiser la façon dont les sujets se *représentent* et évaluent les probabilités. C'est la première apparition explicite de l'idée d'un coefficient de pondération appliqué aux probabilités⁴. Ce facteur est très important pour plusieurs raisons. Tout d'abord, l'idée d'une pondération des probabilités contredit l'idée de N/M (1947), Marschak (1950) et de Samuelson (1952) selon laquelle un agent décide en fonction de probabilités objectives. En fait le sujet décide en fonction de la façon dont il *se représente* les probabilités, comme le dit l'auteur : « De toute façon, il est visible *qu'un individu ne peut tenir compte que des probabilités telles qu'il se les imagine* »⁵. Autrement dit le sujet déforme et distord les probabilités du fait de l'intervention de certains éléments relevant de la représentation, éléments qui seront explicités par la Prospect Theory et analysés plus finement par la

concerne le coefficient de pondération. Les concepts de fonction de valeur et de coefficient de pondération ont fait l'objet de bons nombres d'études, dans le sillage de la Prospect Theory, voir les chapitres 2 et 3 du volume *Choices, Values, and frames*.

¹ Un des buts de notre troisième chapitre sera justement de revisiter ce lien.

² Il faut noter que ces facteurs peuvent agir de façon combinée.

³ Allais (1953), p 508 ; et Allais (1952), p 48-49, où la question est plus élaborée.

⁴ Les remarques dispersées d'Edwards dans son article de 1954 laissent toutefois penser que l'idée était dans l'air. Borel (1941) avait aussi formulé des remarques sur le sujet des représentations des probabilités, notamment en ce qui concerne les probabilités jugées négligeables par le sujet, ou inversement quasi certaines, on trouve aussi des traces de ce problème chez Buffon.

⁵ Allais (1953), *ibid*.

Cumulative Prospect Theory de 1992¹ : il s'agit de la fonction $\pi(p)$ ou coefficient de pondération. Et cette distorsion peut avoir une influence déterminante sur les choix du sujet en provoquant par exemple une aversion² ou un attrait pour le risque qui a pour effet que dans une même situation avec les mêmes enjeux et les mêmes probabilités objectives, deux sujets peuvent décider de façon différente en fonction de leur pondération particulière de l'option. Or cette idée est à l'opposé de ce que proposent N/M dans leur théorie où des sujets différents doivent maximiser de façon identique³.

Par ailleurs, il ne faut pas confondre la déformation subjective des probabilités avec le concept de probabilités subjectives élaboré par Ramsey, de Finetti et Savage. L'axiomatique reste la même pour les probabilités objectives et les probabilités subjectives, et ce n'est pas ce qui est en jeu ici. Cela nous invite à nous demander de quel type de déformation parle Allais, puisqu'il ne mentionne pas une violation des axiomes⁴. Il va s'agir de la façon dont les sujets pondèrent les probabilités au voisinage de la certitude ou de l'impossibilité. Il s'agit en fait d'un véritable changement d'échelle⁵.

Le phénomène de déformation subjective des probabilités est aussi une notion fondamentale concernant les travaux sur les capacités statistiques des sujets en contexte de jugement et de décision que vont mener Kahneman et Tversky avec leur théorie des heuristiques et des biais dans les années 70. Tversky pensait que les humains sont « a priori » de mauvais statisticiens. Les travaux de Cosmides/Tooby (1996) et de Gigerenzer⁶, inspirés

¹ Il s'agit d'une extension de la Prospect Theory qui introduit l'idée d'un dédoublement des coefficients de pondération, voir Kahneman/Tversky (1992), p 60 ; Wakker (2010). Nous reviendrons sur ce problème.

² À ce sujet voir, Pratt (1964) ; Kimball (1993) ; et Rabin/Thaler (2001). L'aversion pour le risque est un concept central en théorie de la décision. Nous l'analyserons exhaustivement avec la Prospect Theory. Dans nos analyses ultérieures nous proposerons une explication originale d'une certaine forme d'akrasie par le biais du concept d'aversion au risque.

³ Il convient ici de se demander dans quelle mesure la théorie de Savage est atteinte par la critique d'Allais ? En fait, cette théorie évite cet écueil de façon provisoire, car les probabilités subjectives que le sujet attribue à un événement peuvent être différentes et par là même produire des fonctions d'utilités différentes pour des individus distincts dans une situation décisionnelle identique. Cependant, dès que nous réintroduisons des éléments « objectifs » dans les probabilités, ce que fait Savage (1954), p 51-55, nous retombons sur le même problème.

⁴ Dans le cas d'un « Dutch book » par exemple c'est un des trois axiomes qui est violé ; voir Skyrms (1986), p 185-189 ; Peterson (2009), p 157.

⁵ C'est ce que va mettre en avant le paradoxe. Voir aussi Granger (1995), p 129-165. Même si du point de vue des probabilités il y a une continuité numérique de 0 à 1, du point de vue de la représentation, les différentes parties de cet intervalle sont évaluées de façons totalement différentes. Il y a en fait plusieurs échelles enchâssées dans cet intervalle que les sujets se représentent souvent par le biais de la notion de « chance de » que nous élucidons dans notre troisième chapitre. La Théorie de la Perspective formalisera comment les probabilités sont distordues, notamment en faisant le graphe du coefficient de pondération ; voir Kahneman/Tversky (1979), p 37. Par ailleurs, cette pondération aura un effet direct sur les préférences des sujets et par là même sur l'axiome d'indépendance.

⁶ Voir, par exemple, Gigerenzer/Hertwig/Pachur (2011). Le propos est de repenser le concept d'heuristique dans une perspective adaptative.

entre autres par l'article de Simon (1956) sur le lien entre comportement rationnel et environnement, viendront modérer le jugement de Tversky, notamment en montrant que les capacités statistiques des sujets sont adaptatives, c'est-à-dire qu'elles s'actualisent et s'ajustent en fonction du contexte.

Le troisième facteur (« élément III ») auquel Allais s'attache est le suivant :

« La pondération suivant les probabilités des valeurs psychologiques et la considération des espérances mathématiques (moment d'ordre 1) de la distribution des probabilités des valeurs monétaires. »¹

Il s'agit ici d'un élément qui dans le processus décisionnel vient synthétiser les deux premiers facteurs : on peut représenter la façon dont le sujet pondère la valeur psychologique d'une perspective en additionnant les options selon la formule suivante :

$$\bar{s}(V) = \bar{p}_1 \bar{s}(\bar{g}_1) + \bar{p}_2 \bar{s}(\bar{g}_2) + \dots \bar{p}_n \bar{s}(\bar{g}_n)$$

Où $\bar{s}(V)$ représente la valeur psychologique de la perspective et $\bar{p}_n \bar{s}(\bar{g}_n)$ représente les options qui la composent et qui font intervenir les éléments I : $\bar{s}(\bar{g}_n)$ et II : \bar{p} ; à savoir la déformation psychologique des valeurs monétaires et la déformation subjective des probabilités objectives. Comme nous le voyons, nous sommes très proches de l'idée de l'espérance morale telle que proposée par Bernoulli (1737) et Laplace (1820). Il ne s'agit pas de combiner de façon stricte les probabilités aux gains comme chez Huygens, mais de combiner ce que le sujet se représente comme un gain probable, en sachant que les probabilités subissent aussi une distorsion, c'est ce qui fait la différence avec la formulation stricte de Bernoulli et de Laplace où les probabilités sont objectives.

Dans son article de 1952², Allais fait remarquer que les éléments I et III peuvent se combiner et produire le phénomène dont nous parlions plus haut, à savoir une perspective

¹ Allais (1953), p 508 ; (1952), p 49.

² p 49-50.

évaluée différemment en fonction des sujets. D'autre part, deux perspectives ayant la même espérance mathématique pourront être évaluées de façons différentes selon la pondération que feront subir les sujets aux options¹.

D'autre part, la possible combinaison des éléments I, II et III amène Allais à remarquer que selon que p est surévalué ou sous évalué, le comportement vis-à-vis des pertes et des gains change². Ce phénomène est important, car il préfigure la principale généralisation de la Prospect Theory³ : la Cumulative Prospect Theory (CPT) construite par Kahneman et Tversky en 1992⁴, dans laquelle les auteurs feront l'hypothèse d'un dédoublement du coefficient de pondération : selon que le sujet est face à une perte ou un gain il évalue les probabilités de façon différente⁵. C'est aussi un point important pour notre travail puisqu'une partie de notre troisième chapitre aura pour vocation de montrer comment le concept de pondération peut être appliqué à d'autres éléments décisionnels que les probabilités⁶.

Le quatrième et dernier facteur (« élément IV ») est pour Allais le plus important : « La prise en considération de la forme des distributions de probabilités des valeurs psychologiques et en particulier leurs dispersions (moment d'ordre 2) »⁷. Ce que signifie cet élément, c'est que les sujets évaluent les perspectives non pas en fonction de la moyenne pondérée des distributions de probabilités, mais en fonction de la *dispersion des distributions de probabilités autour de cette moyenne*. Précisons qu'en prenant en compte les éléments II et IV, le sujet ne se représente pas la valeur de la densité de probabilité⁸ de façon objective, mais de

¹ Mais il peut arriver parfois aussi qu'un même sujet évalue différemment la même perspective selon la forme dans laquelle elle est présentée, c'est ce que vont mettre en avant les études de Kahneman et Tversky avec le concept d'effet de cadrage (« framing effect »). Voir Kahneman/Tversky (1979), p 25-26 ; et (1986). Ces quelques remarques nous introduisent à un problème important de la philosophie de la décision qui est celui de la *singularité du décidable* et du décider, qui fait justement que dans une même situation, deux sujets, voire même un nombre quelconque de sujets, pourront décider de façons totalement différentes. Ce que montrent les travaux d'Allais c'est que la dimension subjective, avec tout ce qu'elle porte, est essentielle dans la théorie du choix rationnel. Il n'y a pas de décision qui ne puisse être ramenée à un système de référence subjectif, nous reviendrons sur ce problème dans notre troisième chapitre.

² Allais (1952), p 51.

³ Dans son texte de 1952, Allais est arrivé très très près de la Théorie de la Perspective, que ce soit au niveau formel ou conceptuel, voir en particulier, p 60-65.

⁴ Voir, Kahneman/Tversky (1992).

⁵ Cette idée se trouve aussi dans Edwards (1954), p 400, nous y reviendrons sous peu.

⁶ En fait le problème est très complexe, il y a deux questions préliminaires auxquelles il faut répondre et qui sont strictement liées : doit-on abandonner la distinction canonique établie par Knight en 1921 entre certitude, incertitude et ignorance ? Par ailleurs doit-on maintenir ou non le postulat de l'utilisation des probabilités par les sujets ? Et si oui, quelles modifications doivent subir les probabilités ?

⁷ Allais (1953), p 509 ; et (1952), p 51 et suivantes.

⁸ Pour ces notions, voir et Baird (1992), p 250-261 ; et Hacking (2004), p 205-267.

façon subjective. Ce facteur est très important, car il rend compte de deux phénomènes assez liés : l'aversion au risque¹ et le plaisir pris au jeu. L'aversion au risque consiste à préférer la valeur d'un pari au pari lui-même. Par exemple si face à une alternative composée d'un pari du type $[(50, .5)(0, .5)]$ et d'une option sûre qui me garantit 25, je préfère l'option sûre, cela signifie que je suis aversif au risque, si ma préférence se porte vers le pari, on pourra alors dire que j'aime le risque et si je suis indifférent entre les deux je serai alors neutre vis-à-vis du risque². Dans le cadre de la théorie d'Allais un sujet aversif au risque préférera une dispersion faible de la distribution des probabilités autour de leur moyenne, un sujet qui aime le risque préférera lui une dispersion importante. Le deuxième phénomène qui trouve une explication grâce à l'introduction du facteur IV est le plaisir pris au jeu. Phénomène dont nous avons vu qu'il n'était pas pris en compte par l'axiomatique N/M : l'axiome 3 : $C : b$ en postulant la possibilité de réduire plusieurs loteries composées à une seule³. Or pour Allais :

« Ce plaisir ou déplaisir attaché au risque vient constituer un élément supplémentaire s'ajoutant au pur calcul fondé sur la pondération probabiliste (objective ou subjective suivant le cas) des satisfactions attachées aux diverses éventualités »⁴.

Une remarque d'ordre un peu plus formel s'impose. À partir du moment où le sujet prend en compte les densités de probabilités, l'évaluation d'une option $s(V)$ doit être fondée sur une fonctionnelle h rapportée à la densité de probabilité $\psi(\gamma)$ des valeurs ψ . Donc :

$$s(V) = h[\psi(\gamma)]^5.$$

¹ Dans son article de 1952, Allais analyse partiellement la relation des sujets au risque.

² Voir, Pratt (1964) ; Kahneman/Tversky (1979) ; Hansson in Gärdenfors/Sahlin (1988), p 136-158 ; Kimball (1993) ; Rabin/Thaler (2001) ; Peterson (2009), p 179-183. Le concept d'aversion au risque sera central pour la Théorie de la Perspective, et nous pouvons dire qu'avec le coefficient de pondération c'est ce qui relie et permet d'établir une continuité entre les travaux d'Allais et la Prospect Theory, il y a cependant des maillons intermédiaires que nous allons étudier. Par ailleurs il faut noter que l'aversion au risque et les phénomènes de pondération sont liés de façon intime : la pondération subjective s'opère à partir de la perception du risque par le sujet.

³ Voir N/M (1947), p 26. Les auteurs font cependant remarquer qu'un autre système est possible si on prend justement en compte ce phénomène, et c'est probablement parce que cet élément a à voir avec la psychologie du sujet que N/M le rejettent de leur axiomatique. Samuelson (1952), sera encore plus explicite par rapport à l'utilisation de cette clause. Pour une analyse claire l'axiome en question voir Luce/Raiffa (1957), p 26-27.

⁴ Allais (1953), p 510.

⁵ Allais (1953) p 513 et (1952), p 55.

Concernant notre propos, il faut remarquer que ce qui est important ici c'est le fait que le sujet oscille autour d'une moyenne. Nous verrons plus tard que cette oscillation constitue un élément central dans le fonctionnement du SRD en ce sens que dans nos existences aussi nous oscillons autour de certains pôles et de certaines habitudes qui peuvent être pensés comme des moyennes. La basse continue de l'existence se situe autour d'une certaine moyenne, moyenne autour de laquelle se déploie notre espace décisionnel. De la même façon que nos sens ne sont pas capables de percevoir des intensités trop fortes ou trop faibles¹, l'existence elle-même se joue à l'intérieur de bornes relativement précises et cela détermine une bonne partie de nos décisions.

Pour Allais ce quatrième facteur est absolument fondamental et selon ses propres mots : on ne peut mener aucune expérience sur le choix rationnel d'où il soit absent. C'est donc : « L'élément spécifique de la psychologie du risque »². Élément qui vient mettre à mal le postulat de la TUA selon lequel on peut construire une fonction d'utilité en observant comment les sujets se comportent face à des situations aléatoires desquelles tout élément psychologique est banni.

2) Paradoxe et rationalité.

Comme nous le remarquons plus haut, Allais distingue deux plans dans lesquels va se déployer sa critique du principe de Bernoulli et par là même de la TUA : la *rationalité abstraite*, qui représente l'aspect normatif de la théorie du choix rationnel et la *rationalité concrète* qui correspond à la dimension descriptive. C'est dans le cadre de l'analyse de la rationalité au sens concret que va se déployer le fameux « paradoxe »³. Nous allons brièvement présenter et discuter les positions d'Allais concernant la rationalité⁴ des agents.

¹ C'est une image que vont aussi utiliser Kahneman et Tversky dans la Prospect Theory, p 32, probablement inspirés par l'utilisation métaphorique de la loi de Weber-Fechner par Allais.

² Allais 1953, p 512.

³ Allais a toujours nié qu'il s'agisse d'un paradoxe.

⁴ La rationalité est un concept clé dans notre travail, notamment en ce qui concerne notre analyse de la TUA, dans laquelle la rationalité est définie par le respect de la structure axiomatique. C'est justement ce qu'Allais va critiquer ici, mais nous verrons qu'il faudra aller encore plus loin lorsque nous étudierons les heuristiques et les biais, la théorie de la perspective et les problèmes liés à la rationalité adaptative. La question à laquelle nous allons nous trouver confronté est celle-ci : qu'est-ce qui doit être modifié de façon profonde, par rapport aux positions de la TUA et de la critique d'Allais, dans la définition du concept de rationalité ? Et ce dans l'hypothèse où le concept de rationalité doit être maintenu. Il s'agira donc d'approfondir et de complexifier le concept de comportement rationnel.

Pour Maurice Allais, la rationalité abstraite peut être définie ainsi : « Un homme est réputé rationnel lorsque (a) il poursuit des fins cohérentes avec elles-mêmes, (b) il emploie des moyens appropriés aux fins poursuivies »¹. Les clauses (a) et (b) ont pour corollaires que 1) le sujet doit pouvoir ordonner les alternatives ou les options², 2) qu'il doit choisir l'alternative qui domine l'autre sous tous les aspects, c'est-à-dire dont le gain est le plus élevé dans tous les cas. Allais baptise cet axiome : « axiome de préférence absolue »³. Le troisième postulat qu'Allais « déduit » des éléments (a) et (b), est l'utilisation des probabilités objectives par le sujet. Ces trois postulats suffisent⁴ à définir la rationalité d'un point de vue abstrait, c'est-à-dire normatif.

À la suite de l'introduction de ces clauses l'auteur remarque que ni le P2 de Savage ni le principe d'indépendance de Samuelson ne peuvent entrer dans la définition de la rationalité, car ils négligent tous deux l'élément IV, que nous avons analysé plus haut, à savoir la dispersion des valeurs psychologiques et l'ampleur que peut avoir cette dispersion pour un sujet. À vrai dire on ne voit pas très bien comment relier la critique de Savage et de Samuelson à ce qui vient d'être dit sur la rationalité abstraite ; ce que critique Allais ici, c'est en fait la visée normative de ces axiomes. Par exemple l'axiome P2 de Savage ne peut être normatif, car il stipule qu'un changement dans les composantes communes aux options ne doit pas altérer les préférences des sujets, or dans le cadre de la théorie d'Allais cette modification⁵ doit avoir un impact sur les dispersions des valeurs psychologiques et donc sur les préférences : le sujet n'oscille plus autour de la même densité de probabilités.

Avant de passer à la réfutation de la TUA via celle du principe de Bernoulli, Allais donne une première définition de la rationalité expérimentale : « Ce que font effectivement les

¹ Allais (1953), p 518 ; et (1952), p 69. L'auteur précise que sa définition est élaborée à partir de « la logique scientifique », et nous ne voyons pas très bien à quoi il fait référence ici. S'agit-il de la logique entendue comme théorie normative du raisonnement valide ? Dans ce cas Allais se rapprocherait de Savage et de Ramsey. En tout cas il y a un aspect normatif évident dans la définition que l'économiste propose, définition qui comme le souligne Amihud (1979), p 150, a tout de même de sérieuses faiblesses.

² Ce qui signifie qu'Allais utilise quand même un axiome d'ordonnancement, qu'il ne détaillera pas plus d'ailleurs. Et remarquons qu'Allais ne définit pas, comme il le note p 522, la rationalité par l'ordonnancement ou par son axiome de préférence absolue.

³ Allais (1953), *ibid.* Cet axiome est analysé de façon exhaustive, avec schéma à l'appui dans le texte de 1952 aux pages 39 à 41. Le but de cet axiome est de garantir le respect de la dominance stochastique d'ordre 1. À savoir que le sujet doit choisir l'option qui domine l'autre pour toutes les conséquences possibles. Voir Machina (1987), p 127-132 ; Willinger (1990), p 11-15.

⁴ Comme le remarque l'auteur ils sont « les seules conséquences de la condition de cohérence », p 519.

⁵ Ainsi que le note Machina (1987), p 129 ; et Willinger (1990), p 14-15, cette modification a une influence sur la dominance stochastique des options.

hommes dont on a par ailleurs des raisons¹ de penser qu'ils se comportent rationnellement »². Ce qui nous intéresse dans le cadre de ce travail c'est justement la dimension expérimentale, car elle va ouvrir la voie à une série d'études – venant de divers domaines de recherche (économie, psychologie, mais aussi philosophie) – dont la finalité sera l'analyse systématique des violations empiriques de la TUA. Analyses dont les résultats nous fourniront les éléments nécessaires à la constitution de nos propres hypothèses. D'autre part, c'est dans le cadre de la réfutation expérimentale de la TUA, et en particulier des critiques concernant l'axiome d'indépendance³, qu'Allais va introduire son « paradoxe ».

Toutefois, nous allons dire quelques mots rapidement de la réfutation qui se situe dans la dimension de la rationalité abstraite. Il va s'agir pour l'économiste de montrer deux choses : que son concept de rationalité *n'implique pas* l'idée de maximisation, mais seulement les quatre éléments que nous venons d'étudier. D'autre part, Allais entend prouver que pour différents types de choix rationnels, la formulation de Bernoulli est trop restrictive ou invérifiable⁴.

Comme nous le remarquons, ici Allais se place au point de vue logique et se pose cette question : ma définition de la rationalité abstraite⁵ implique-t-elle l'hypothèse de maximisation ? Or, dans les trois exemples de choix entre des perspectives aléatoires qu'Allais propose, l'implication de fonctionne pas.

Il va s'agir maintenant pour l'économiste de raffiner sa définition de la rationalité concrète en spécifiant que nous pouvons nous faire une idée de ce type de rationalité par : « l'observation de ce que font des hommes rationnels et tout à fait au courant du calcul des probabilités »⁶. À partir de cette définition, deux éléments à charge contre la TUA vont

¹ Mais quelles sont ces raisons ? Allais ne les précisent jamais, ce qui constitue, du point de vue philosophique une lacune assez grave. Encore une fois cette définition approximative sera critiquée par Amihud (1979), p 150 et suivantes.

² Allais (1953), p 521.

³ Il faut garder à l'esprit que cet axiome est le cœur de la TUA, et que c'est la raison pour laquelle il sera la cible privilégiée des attaques. L'idée étant que si l'axiome est réfuté empiriquement, du point de vue normatif et descriptif, alors c'est toute la TUA qui s'effondre. Machina (1987), p 124-132 ; et Willinger (1990), p 8-15, proposent une analyse très claire des problèmes que soulève la clause d'indépendance, à la fois par rapport à la linéarité vis-à-vis des probabilités, et par rapport à l'idée de maximisation. Nous renvoyons le lecteur à notre analyse de l'axiome dans la partie précédente pour plus de références et à l'article central de Fishburn/Wakker (1995).

⁴ Allais (1953), p 523-524.

⁵ Il est bien entendu que l'auteur refuse catégoriquement la définition des promoteurs de la TUA, à savoir que la rationalité des sujets consiste dans le respect de la cohérence axiomatique.

⁶ Allais (1953), p 524 ; (1952), p 79. Encore une fois, remarquons le caractère vague et imprécis de cette définition : de quel type d'observation s'agit-il ? Quotidienne, ou scientifique, protocolaire ? Probablement, Allais a-t-il à l'idée l'expérience qu'il avait menée au Congrès de 1951 à Paris, et où il avait piégé Savage, entre autres. Mais quoi qu'il en soit sa définition n'est pas construite du point de vue philosophique et elle fait intervenir le *definiendum* dans le *definiens*.

émerger : le comportement des sujets face à des loteries dont les lots sont des petites sommes et la mise en échec du P2 de Savage dans les choix aléatoires dont les probabilités se situent non loin de la certitude ou de l'impossibilité¹.

Le premier phénomène fait intervenir l'aversion au risque des agents, c'est-à-dire l'effet combiné des facteurs I et IV² que nous avons analysé plus haut. En effet, dans le cadre d'un choix unique il est tout à fait rationnel de préférer 40 \$ sûrs à ½ de gagner 100\$, ou 400 sûrs à 1000 avec une chance sur 2. Il n'y a *aucune raison contraignante*, si ce n'est le respect de l'hypothèse de maximisation, de considérer de tels choix comme irrationnels : il s'agit simplement de sujets qui font preuve d'une aversion pour le risque, autrement dit de sujets que nous pouvons qualifier comme « prudents »³. Le fait que l'utilité attendue de la perspective incertaine soit plus élevée n'entre pas en ligne de compte. Allais critique l'idée de maximisation, les sujets ne maximisent pas⁴, et comme le note l'auteur : « Ils ont une préférence marquée pour la sécurité »⁵.

C'est la critique de l'axiome d'indépendance qui va conduire Allais à formuler le fameux « paradoxe ». Paradoxe qui est issu d'une expérience que l'économiste avait menée lors d'un congrès sur la décision en contexte d'incertitude à Paris en 1951. Savage été présent, et a été piégé par le questionnaire d'Allais, avec quelques autres représentants de la TUA⁶.

Il va s'agir pour Allais de montrer d'une part que les sujets ne maximisent pas et d'autre part qu'ils ne respectent pas la propriété d'indépendance par rapport aux probabilités⁷. Afin de parvenir à son but, Allais va faire intervenir le concept central de certitude et son influence dans toute situation décisionnelle. Ce que l'économiste entend montrer c'est que les

¹ Rappelons une fois encore que le P2 de Savage est une clause d'indépendance, voir Savage (1954), p 23-24. En ce qui concerne les probabilités utilisées dans les exemples d'Allais, Morgenstern (1979), leur reprochera leur caractère peu réaliste, il en ira de même pour les sommes engagées.

² Dans un souci plus historique, il serait intéressant de retracer la genèse de ces éléments, d'en repérer l'émergence dans l'histoire de l'économie, étude qui n'a pas été encore menée avec toute la rigueur souhaitée. On peut néanmoins trouver des éléments dans Stigler (1951).

³ Et on sait le rôle de premier plan que tient la prudence dans les processus délibératifs, comme en témoignent les réflexions d'Aristote au Livre VI de *L'Éthique à Nicomaque* concernant le lien entre le *phrônimos* (prudent, ou sagace, si l'on préfère la traduction de Bodéüs) et la décision, p 302 et suivantes. À ce sujet voir aussi le classique de P. Aubenque (1963).

⁴ Comme le remarque Edwards (1954), à partir du début des années 50, la TUA et l'idée de maximisation seront soumises à des tests expérimentaux de plus en plus élaborés, voir par exemple Mosteller/Nogee (1951), qui concentrent une partie leurs expériences sur le concept de transitivité.

⁵ Allais (1953), p 525. C'est un élément important qu'il faudra repenser en connexion avec le système de référence décisionnel.

⁶ Voir Savage (1954), p 101-104.

⁷ Voir Machina (1987) ; Willinger (1990).

sujets décident en fonction de la proximité ou de l'éloignement par rapport à la certitude, et que des gains certains ont toujours un attrait plus grand. Le comportement des décideurs face aux mêmes alternatives sera différent selon que les probabilités se situent proches de la certitude ou y sont égales, ou selon que les probabilités en sont plus ou moins éloignées. Il s'agit ici de mettre en relation le probable et le possible et de se demander dans quelle mesure le probable peut être la mesure du possible, et à quel moment le probable et le possible se déconnectent l'un de l'autre¹. À partir de là émergent des éléments critiques pour la TUA et fondamentaux pour la suite de nos recherches : l'effet de rapport commun, l'effet de conséquence commune, et le rôle précis joué par l'aversion au risque.

Comme nous le remarquons dans notre partie précédente le cœur de la TUA est l'axiome d'indépendance qui a pour rôle de garantir l'indépendance des préférences par rapport aux croyances exprimées en termes de probabilités objectives ou subjectives². Comme le remarquait Fishburn (1973), sans cet axiome l'idée de maximisation de l'utilité attendue disparaîtrait de la TUA³. Afin de mettre en question la clause d'indépendance de façon empirique, Allais propose le choix suivant⁴ :

« (1) *Préférez-vous la situation A à la situation B ?*

Situation A : Certitude de recevoir 100 millions.

Situation B : 10 chances sur 100 de gagner 500 millions.

¹ On peut par exemple postuler qu'à partir d'un certain seuil, des probabilités minimales sont assimilées à l'impossibilité. C'est ce qu'avait déjà évalué Borel dans son ouvrage de 1944. Pour une analyse plus poussée de ces problèmes, du point de vue de la philosophie, voir Largeault (1979) ; et surtout Granger (1995).

² Voir Machina (1987), p 123-127. Cet article est une synthèse très éclairante des travaux qui ont vu le jour en théorie de la décision à partir des contestations empiriques de Allais (1953) et Ellsberg (1961). Voir aussi Willinger (1990), p 9, qui s'inspire d'ailleurs de Machina.

³ Fishburn (1973), p 108.

⁴ Le « paradoxe » a donné lieu à une littérature plus qu'abondante. Il n'existe pas à notre connaissance d'ouvrage en théorie de la décision qui ne le mentionne et ne lui réserve une importante partie. Jusqu'en 1979, l'ouvrage de référence reste celui coordonné par Allais et Hagen : *Expected Utility Hypotheses and the Allais Paradox*. C'est un ouvrage important, car il contient l'article d'Allais de 1952 qui est beaucoup plus long et explicite que celui de 1953, mais ce livre contient aussi des critiques pertinentes venant par exemple, entre autres, d'Oskar Morgenstern, il contient aussi la réaction et les réflexions d'auteurs comme Machina ou Fishburn à propos du paradoxe. La même année, Kahneman et Tversky vont publier leur article séminal sur la Prospect Theory, largement inspiré par Allais (1953). D'autres éléments intéressants quoique pas toujours directement liés au paradoxe sont à chercher dans un autre livre coordonné par Allais et Hagen : *Cardinalism* (1994). Il n'y a pas eu à proprement parler de livre de synthèse depuis celui de 1979, mais on peut trouver des éléments bibliographiques intéressants dans Baron (1988) ; Gärdenfors/Sahlin (1988) ; Barberà/Hammond/Seidl (2004) ; Peterson (2009) ; Binmore (2009) ; Gilboa (2009). Les articles qui analysent ou discutent le paradoxe sont légions, parmi cette masse cinq ont retenu notre attention : Machina (1987) et Willinger (1990), pour leur aspect synthétique ; Birnbaum (2004) pour l'analyse des rapports entre le paradoxe et les différentes théories de l'utilité qui ont vu le jour après sa publication ; Hussain/Saz/Zafar Yaqub (2009) qui est une revue de littérature très complète sur le paradoxe ; et enfin Birnbaum/Schmidt (2010), pour le rapport entre le paradoxe et les effets de cadrage.

89 chances sur 100 de gagner 100 millions.

1 chance sur 100 de ne rien gagner.

(2) *Préférez-vous la situation C à la situation D ?*

Situation C : 11 chances sur 100 de gagner 100 millions.

89 chances sur 100 de ne rien gagner.

Situation D : 10 chances sur 100 de gagner 500 millions.

90 chances sur 100 de ne rien gagner. »¹

Dans le cadre de la TUA, l'axiome d'indépendance nous permet de prédire que les choix rationnels sont les suivants : si $A \succ B$ alors $C \succ D$, justement du fait de la propriété d'indépendance des préférences par rapport aux probabilités². Or, il fut observé que la plupart du temps $A \succ B$ et $C \prec D$ ce qui vient immédiatement contredire le pouvoir *normatif* et *prédictif* de l'axiome d'indépendance³.

Il faut à présent regarder les choses d'un peu plus près afin de donner une analyse exhaustive de ces résultats. Tout d'abord nous assistons ici à deux violations de la TUA : violation de la propriété d'indépendance et violation de l'hypothèse de maximisation. Calculons l'utilité attendue des quatre options : $U(A)=100\bar{M}$, $U(B)=139\bar{M}$, $U(C)=11\bar{M}$, $U(D)=50\bar{M}$. Or, dans la situation (1) les sujets se portent massivement vers une option qui contredit de façon immédiate l'idée de maximisation⁴ :

$$U(100\bar{M}) \succ U(139\bar{M})$$

¹ Allais (1953) p 527 et (1952) p 89.

² Dans leurs expériences de 1979, Kahneman et Tversky, reprenant exactement la même structure que celle du « paradoxe », montreront que 80% des sujets ont des préférences qui violent la propriété d'indépendance ; voir Kahneman et Tversky (1979), p 20-21.

³ Voir Kreps (1988), p 192-194 ; Picavet (1996), p 229-239 ; Peterson (2009), p 78-81. En effet, ces préférences démontrent que les probabilités ont une influence très nette sur le choix des sujets. Comme le remarque Allais, et ce du fait de l'intervention des facteurs II et IV, c'est le passage de la certitude à la possibilité qui influence ici les sujets ; voir Allais (1953), p 528.

⁴ Cela signifie que la plupart des sujets ne calculent pas. C'est un fait important qu'il conviendra d'analyser en profondeur. En effet, une tradition, notamment kantienne, associe aux impératifs hypothétiques de la prudence une forme de calcul : si tu veux telle fin, choisis tel moyen, voir *Grundlegung zur Metaphysik der Sitten*. Zweite Abschnitt. Tradition qui s'origine en partie dans l'éthique aristotélicienne, éthique qui conçoit elle aussi le processus délibératif comme une forme de calcul en vue de trouver des moyens d'atteindre une fin déterminée, voir *Éthique à Nicomaque*, III, 1122 b18-23. Pour une comparaison entre les deux points de vue, voir Aubenque (1963), p 186-221. Il conviendra dans notre dernier chapitre de reprendre cette interrogation, en nous demandant notamment, si nous pouvons vraiment parler de calcul quand nous analyserons l'aspect évaluatif et pondérateur des processus volitifs.

D'autre part un simple calcul nous permet de montrer qu'il n'existe pas de fonction d'utilité qui puisse décrire simultanément $A \succ B$ et $C \prec D$ car si nous calculons les différences d'utilité attendue entre les deux perspectives, nous découvrons ceci :

$$\begin{aligned} U(A) - U(B) &= U(100\bar{M}) - [0.01U(0\bar{M}) + 0.1U(500\bar{M}) + 0.89U(1\bar{M})] \\ &= 0.11U(1\bar{M}) - [0.01U(0\bar{M}) + 0.1U(500\bar{M})] \end{aligned}$$

$$\begin{aligned} U(C) - U(D) &= [0.11U(100\bar{M}) + 0.89U(0\bar{M})] - [0.9U(0\bar{M}) + 0.1U(500\bar{M})] \\ &= 0.11U(100\bar{M}) - [0.01U(0\bar{M}) + 0.1U(500\bar{M})] \end{aligned}$$

La différence d'utilité attendue entre les deux perspectives est la même, ce qui signifie que du point de vue de la TUA il est tout simplement impossible de préférer A à B et D à C . C'est pourtant la violation de la propriété d'indépendance qui doit être considérée de façon plus approfondie. L'axiome d'indépendance, par exemple dans la forme que lui donne Savage, nous indique que les conséquences communes aux options ne doivent pas être prises en compte dans les décisions. Or, le paradoxe d'Allais met en avant le phénomène inverse qui sera baptisé « effet de la conséquence commune » (« common consequence effect »)¹. Dans la présentation que donne Allais du paradoxe, la conséquence commune n'apparaît pas de façon très claire. Nous allons utiliser le principe de la formulation de Savage², c'est-à-dire présenter les événements sous forme de tickets de loteries numérotés de **1** à **100**, et les **situations I et II** comme des paris dont **A/B**, **C/D** sont les alternatives. Ce qui donne le tableau suivant :

¹ Cette appellation vient du fait que les options A et B d'une part, C et D , d'autre part, partagent une conséquence commune. Voir Machina (1987), p 129, qui formalise ceci de façon très claire. Ce phénomène a aussi fait l'objet d'études intéressantes de la part de Birnbaum (2004) qui propose une véritable « dissection du paradoxe d'Allais », notamment en utilisant certains concepts issus de la Théorie de la Perspective. Cet article montre que le paradoxe peut être évité si la TUA est modifiée par l'introduction de certaines clauses comme la coalescence ou une version affaiblie du PCS (« Restricted Branch Independence »). La question implicite sur laquelle fait fond l'article, à savoir comment construire une théorie qui puisse se passer de l'axiome d'indépendance, ou tout au moins utiliser une version plus faible de cet axiome, est magistralement étudié par Machina (1987), p 132-150 ; voir aussi Willinger (1990), p 18-43 ; Picavet (1996) aborde le problème aux pages 228-251 ; enfin, on peut se reporter à la recension de R. Sugden in Barberà/Hammond/Seidl (2004), p 730-749, et à celle d'U. Schmidt dans le même volume, aux pages 758-837, concernant les modifications formelles qu'il faut apporter à la TUA si on veut se passer ou affaiblir l'axiome d'indépendance. Il faut toutefois remarquer que ce qui fait l'unité de ces tentatives théoriques, c'est la volonté de rester aussi proche que possible du modèle N/M en faisant varier les axiomes en fonction des contestations empiriques, mais en maintenant toutefois l'idée de maximisation. Un bon exemple de ce type de théorie est la Théorie du regret de Loomes/Sugden (1982).

² Savage (1954), p 103.

		1	2-11	12-100
Situation I	A	$100\bar{M}$	$100\bar{M}$	$100\bar{M}$
	B	$0\bar{M}$	$500\bar{M}$	$100\bar{M}$
Situation II	C	$100\bar{M}$	$100\bar{M}$	$0\bar{M}$
	D	$0\bar{M}$	$500\bar{M}$	$0\bar{M}$

Il convient ici de remarquer deux choses. Dans le cas où l'événement [12-100] se réalise, les options A/B et C/D ont les mêmes conséquences. *Conséquences qui ne devraient donc pas entrer en ligne de compte si le PCS était respecté.* Deuxième élément important, qui rejoint ce que nous avons vérifié par le calcul : si nous faisons abstraction des conséquences communes les **situations I et II** sont identiques. Interrogeons-nous alors sur ce qui motive les préférences exprimées par les sujets.

En fait il y a plusieurs facteurs qui interviennent ici. Tout d'abord il y a *l'option certaine A* qui a une influence particulière : dans quelle mesure peut-on considérer comme irrationnel un sujet qui préfère $100\bar{M}$ sûrs à 0.98 de gagner $500\bar{M}$ ¹? C'est d'ailleurs ce phénomène qui cause aussi la violation du principe de substituabilité de Samuelson², principe qui est l'explicitation affaiblie de la propriété d'indépendance de N/M³. Les clauses d'indépendance de ces auteurs nous indiquent qu'il ne peut y avoir aucun type de complémentarité entre les options. C'est justement ce qu'Allais veut réfuter ici puisqu'il « suffit de rechercher des cas où P(3)⁴ aura avec P(1) et P(2) des effets différents de complémentarité (ou de non-complémentarité) qui puissent changer l'ordre des préférences »⁵.

En ce qui concerne l'option A, l'économiste a eu ici l'intuition de ce qui allait devenir *l'effet certitude* (« certainty effect ») en Théorie de la Perspective, effet qui doit être connecté avec le concept d'aversion au risque⁶. Allais a anticipé sur bien des points la Théorie de la perspective, surtout dans son article de 1952. Remarquons que c'est à partir d'une analyse de l'aversion au risque et de l'effet certitude que nous proposerons l'explication d'un certain type d'akrasie. Dans ce cadre-là, un choix akratique serait dû à une pondération biaisée de l'option

¹ Allais 1953, p 528.

² Voir Samuelson (1952).

³ Voir Malinvaud (1952).

⁴ Il s'agit des probabilités des options qu'Allais utilise pour présenter l'axiome de Samuelson.

⁵ Allais, ibid.

⁶ Voir Kahneman/Tversky (1979), p 20-22.

par rapport au statu quo, ou inversement d'une surpondération du statu quo par rapport à une option plus intéressante. Le risque serait de *perdre* quelque chose, un état, une habitude qui me satisfait, et dont de je n'arrive pas à évaluer si la privation ou la modification m'apporterait quelque chose de plus, c'est-à-dire me ferait *gagner* quelque chose.

Revenons au paradoxe, et à sa présentation sous forme de tableau qui permet de faire apparaître de façon plus claire les conséquences communes. Comme nous le remarquons, si le PCS est respecté, ces conséquences ne doivent avoir aucune influence sur le choix des sujets. Cependant lorsque nous passons de **SI** à **SII** nous remarquons que les conséquences communes de **SII** ($0\bar{M}$) sont moins intéressantes que celles de **SI** ($100\bar{M}$). Ce qui signifie que **SI** domine stochastiquement **SII**¹.

Le paradoxe ferait donc apparaître que les sujets sont sensibles à la dominance stochastique dont ne parvient pas à rendre compte le principe d'indépendance. Dans le cadre de notre tableau les sujets seront sensibles à l'alternative qui est la plus intéressante du point de vue de la dominance stochastique et préféreront donc minimiser le risque dans ce cas-là. Or la dominance stochastique s'exprime justement dans les options communes que les sujets sont censés ignorer. D'où le choix de **A** dans **SI**, qui est un choix où la variance est nulle, par rapport à **B**, où la variance, bien que minime peut faire « louper » au sujet le gain de $100\bar{M}$. Ce sera l'inverse qui motivera les décideurs dans la **situation II** qui est donc stochastiquement dominée par la **situation I**. Dans le cas de cette alternative, la conséquence commune est très faible mais elle vient avec une grande probabilité. Ici, les sujets vont choisir l'option qui a l'utilité attendue la plus élevée contrairement à ce qui se passe dans la première situation. La dominance stochastique est donc d'une certaine façon liée avec l'aversion au risque : le choix pour l'option **A** dans la **situation I** montre que les agents sont ici très aversifs au risque, alors que le choix de **D** dans la **situation II** montre justement l'inverse, dans ce cas de figure là les sujets aiment le risque. La prise en compte de la dominance stochastique par les sujets produit un renversement des préférences².

¹ Ici il s'agit de dominance stochastique d'ordre 1 qui signifie que toutes les conséquences de **SI** sont au moins aussi bonnes ou meilleures que celle de **SII**. Formellement, si x est une conséquence quelconque, et si $F(x)$, $G(x)$, sont des actes qui donnent x si un événement de références est le cas, on dira que F domine stochastiquement G si et seulement si $(\forall x)[F(x) \geq G(x)]$.

² Voir Machina (1987), p 129 ; Willinger (1990), p 14. Le concept de *renversement des préférences* (« preference reversal ») sera analysé sous peu, à partir des travaux de Lichtenstein et Slovic (1968), (1971), (1973) et (2006). Encore une fois nous

Allais voulait aussi montrer dans son article à quelle condition et dans quel contexte très limité et délimité l'hypothèse de maximisation fonctionne : *seulement lorsque l'utilité attendue est équivalente à la valeur psychologique de la perspective*, ce qui est le cas dans la **situation II**. Car dans la **situation I**, l'option certaine produit un véritablement changement dans l'échelle d'évaluation du sujet, due entre autres à l'intervention de l'élément II, et par là même une totale dissociation entre l'utilité attendue de la perspective et sa valeur psychologique.

Un autre effet que va mettre en avant Allais dans son article de 1953 – qui vise cette fois plus spécifiquement l'axiome de N/M – est l'effet du *rapport commun* (« common ratio effect »)¹ qui réfute l'indépendance des préférences par rapport aux probabilités². L'exemple que nous allons étudier est établi à partir de la critique de l'axiome de substituabilité de Samuelson³. Pour simplifier, imaginons que l'on présente aux sujets un choix entre les options F et G , F' et G' :

F : 100 millions sûrs.

F' : 0.01 de gagner 100 millions.

0.99 de gagner 1 franc.

G : 0.98 de gagner 500 millions.

G' : 0.0098 de gagner 500 millions.

0.2 de ne rien gagner.

0.99 de gagner 1 franc.

0.0002 de ne rien gagner.

Encore une fois si le principe d'indépendance de N/M était respecté un choix tel que $F \succ G$ devrait nous permettre de prédire que $F' \succ G'$. Or, en général les sujets préfèrent G' à F' . En fait si nous regardons les choses de façon un peu plus formelle nous nous rendons compte que les rapports entre les probabilités des gains restent constants (d'où le nom de ce phénomène) : il y a le même rapport entre 1 et 0.01 qu'entre 0.98 et 0.0098, ce qui devrait conduire à des préférences constantes. Or cette identité de rapports masque en fait un

pouvons dire que ce concept était dans l'air du temps, le paradoxe d'Allais permet aussi de mettre en avant ce genre phénomène, à cette différence près toutefois que ce que vise l'économiste ce n'est pas l'ordonnancement et la transitivité des préférences.

¹ Il a aussi été étudié par Hagen (1979) ; Kahneman/Tversky (1979) ; Prelec (2000), p 71.

² Voir Machina (1987), p 130 ; Koehler/Harvey (2004), p 403 ; Barberà/Hammond/Seidl (2004), p 694-699.

³ Cette critique occupe les pages 529-530 de l'article. Il est important de garder à l'esprit que c'est au travers l'axiome de Samuelson qu'est explicitée, certes d'une façon un peu affaiblie, la propriété d'indépendance de N/M. C'est ce que démontre l'article de Malinvaud (1952). C'est donc en particulier N/M qui sont visés au travers de l'exemple que nous allons étudier.

changement d'échelle dû à l'évaluation des probabilités :

« Ces choix correspondent en réalité à des situations complètement différentes et soutenir que, dans ces situations tout à fait différentes, il faut du point de vue rationnel se comporter de la même manière, c'est énoncer un postulat qui, lui n'a rien d'évident ni même d'attractif, au contraire du postulat de substituabilité »¹

D'autre part, dans les deux options, le rapport de $\frac{p(500)}{p(100)} = 0.98$. Encore une fois, cette égalité n'est pas prise en compte et conduit à un ordonnancement des préférences violant l'axiome d'indépendance. D'un point de vue plus formel et en simplifiant les alternatives à deux options, nous avons les préférences suivantes :

$$\begin{aligned} (p, x ; 1-p, y) &> (q, x ; 1-q, y) \\ &\text{et} \\ (rp, x ; 1-p, y) &< (rq, x ; 1-rq, y) \\ &\text{or} \\ \frac{p(x)}{rp(x)} &= \frac{q(x)}{rq(x)} \end{aligned}$$

Que ce soit dans le cas de l'effet de conséquence commune ou dans le cas de l'effet de rapport commun les sujets n'évaluent donc pas les perspectives et ne les comparent pas selon leurs valeurs réelles du point de vue des probabilités et de l'utilité. Il y a des éléments qui sont à verser au registre de la *représentation subjective* qui intervient et déforme la valeur réelle des options.

Cependant comme le remarque Allais, de tels sujets ne peuvent être considérés comme irrationnels : « parce qu'ils attachent une grande valeur à la certitude, et que loin de la certitude, ils pondèrent les valeurs psychologiques suivant les probabilités »². Ce qui indique au fond que les sujets sont sensibles à la dispersion des valeurs psychologique, c'est-à-dire à l'élément IV, élément central que l'auteur avait mis en avant lors de son analyse préparatoire du choix rationnel en contexte de risque. Ce qui le conduit à cette conclusion :

¹ Allais (1953), p 540.

² Ibid.

« Il résulte de tout ce qui précède que l'erreur fondamentale de toute l'école américaine, c'est de négliger indirectement et inconsciemment, la dispersion des valeurs psychologiques »¹.

Nous concluons cette analyse par quelques remarques de différentes sortes concernant la portée de cet article qui fit et fait encore couler beaucoup d'encre. Tout d'abord, l'article de 1953 eut un impact immédiat sur la théorie néo-bernoullienne de l'utilité. Après Allais, la question devenait celle-ci : comment adapter la TUA en prenant en compte les contestations empiriques de l'axiome d'indépendance? Et la critique du PCS par Ellsberg (1961) allait encore aggraver la situation ².

Cependant, il ne faudrait pas exagérer l'impact « destructeur » de la critique d'Allais. En effet, dans son article, l'économiste met certes en cause la clause d'indépendance, qui est cruciale, et aussi celle portant sur les rapports entre le plaisir pris au jeu et la réduction des loteries composées, mais, comme le note O. Morgenstern (1979) cette critique ne remet jamais en question la cohérence logique interne de l'axiomatique N/M³. Ce que note aussi Morgenstern, c'est que pour des probabilités qui ne sont pas extrêmes et des enjeux « normaux », la TUA reste la meilleure théorie qui soit à la fois descriptive et normative⁴. Cette théorie reste applicable pour des alternatives qui se situent dans des limites raisonnables.

D'ailleurs, la réaction de Savage au paradoxe mérite d'être soulignée encore une fois : sa réponse prouve que la TUA fonctionne du point de vue normatif pourvu que nous nous donnions la peine de reconstruire le paradoxe et de l'analyser à un autre point de vue et dans une autre forme⁵. Le découpage de la réalité décisionnelle que propose Savage permet

¹ Allais (1953), p 544. Remarquons que l'auteur ne va pas plus loin en ce qui concerne l'intervention de facteurs inconscients dans l'erreur de l'école américaine. Et nous pouvons nous demander à juste titre ce qu'il entend par là.

² Voir Machina (1987), p 132-136. L'auteur parle de modèles « non-expectatifs », c'est-à-dire de modèles qui font intervenir un certain nombre de coefficients pouvant pondérer par exemple les probabilités ou les valeurs des issues. Cependant la formulation reste très proche de celle de N/M, et à cet égard nous aurons l'occasion de nous rendre compte que la Théorie de la perspective est très conservatrice. Voir Kahneman/Tversky (1979). Les impacts du paradoxes d'Allais sur la théorie de l'utilité sont analysés

³ Voir Morgenstern, in Allais/Hagen (1979), p 175-183. L'auteur insiste aussi sur l'aspect non-réaliste des alternatives dans le paradoxe qui engagent des sommes énormes avec des probabilités extrêmes.

⁴ Ce qu'avaient confirmé les expériences de Mosteller/Nogee (1951).

⁵ Si on reconstruit le paradoxe comme le fait Savage (1954), p 103, il appert que les incohérences originelles sont à verser au registre d'un effet de cadrage, c'est-à-dire à la façon dont la présentation des options influence le choix des sujets.

d'absorber le paradoxe¹. Si nous nous plaçons du point de vue strictement logique, force est de constater que la théorie de Savage est bien plus solide que celle d'Allais et qu'elle laisse place à moins de flottement en ce qui concerne les définitions. Seulement ce sont deux visions de la décision humaine qui s'affrontent ici, et il demeure vrai que les exigences de rationalité de Savage sont exorbitantes pour des sujets dotés d'une rationalité limitée, alors que l'approche d'Allais semble plus réaliste et proche de l'expérience. Ce qu'Allais gagne en réalisme, il le perd du point de vue de la cohérence logique et de la force mathématique, nous aurons l'occasion de revenir sur le concept de cohérence et sur la place problématique qu'il occupe dans le domaine décisionnel.

Une critique qui peut être adressée à Allais, c'est la faiblesse de sa définition de la rationalité, que ce soit dans le domaine abstrait ou concret. Il eut peut-être été intéressant et important de construire ce concept d'une façon plus solide et systématique², d'une façon plus philosophique. Les deux définitions qu'Allais donne, et notamment, comme le note Amihud (1979)³, celle qui fait appel « à des gens jugés comme parfaitement rationnels » sont insuffisantes : en effet qui juge que ces gens sont rationnels et d'après quels critères ? C'est une question capitale à laquelle Allais ne répond pas.

Dernier point, pour nous le plus important : la façon dont l'article d'Allais anticipe sur bien des points la théorie de la perspective en faisant intervenir le concept de pondération psychologique des valeurs et des probabilités. Pondération qui comme nous le remarquons plus haut est à verser au registre des effets de la façon dont *les sujets se représentent les alternatives*. Comme nous nous en rendrons compte dans la TP et d'une façon plus philosophique dans notre troisième chapitre, c'est dans l'*édition* des perspectives que se joue une grande partie de la prise de décision, et c'est probablement là aussi que se fabriquent les « paradoxes » comme celui d'Allais ou celui d'Ellsberg. Or cette phase d'édition est complètement conditionnée par un ensemble de représentations subjectives que nous détaillerons plus tard. Cette entrée en scène du concept de représentation, certes encore vague et à constituer philosophiquement, marque un moment important de notre travail : une

¹ Savage (1954), p 103. Voir aussi les remarques de Lichtenstein/Slovic (2006), p 34-35. Les auteurs utilisent le terme « reframe » pour désigner la reconstruction que fait Savage du paradoxe.

² Comme nous le notions plus haut, la TUA est à cet égard bien plus solide, quoique très abstraite du point de vue du comportement réel des sujets.

³ P 150.

première « incarnation » psychologique du sujet en contexte décisionnel, incarnation qui ira en s'approfondissant dans la suite de notre étude.

L'article d'Allais et les travaux qui viendront le compléter du point de vue de la psychologie expérimentale représentent le passage d'un agent rationnel conçu comme un décideur idéal – dans la TUA – à un sujet réel confronté à des décisions réelles¹, c'est-à-dire ayant comme toile de fond, d'une part le monde comme réalité concrète soumise à des processus dynamiques et historiques, d'autre part son propre système de référence général et décisionnel. Il conviendra dans notre troisième chapitre de donner à ce sujet toute sa densité phénoménologique, notamment en réinjectant le concept de volonté dans le domaine de la décision.

¹ Avec toutefois les réserves qui s'imposent lorsque nous utilisons ce mot dans un contexte expérimental.

B – L’approfondissement de la critique : *probabilités, contexte et ordonnancement*.

1) La synthèse critique de Ward Edwards.

L’article de Ward Edwards¹ : « The Theory of Decision making » publié en 1954² est une synthèse – doublée d’une revue de littérature – dont l’objectif est de mettre en relation les théories de la décision en contexte de certitude et en contexte de risque avec leurs confirmations ou contestations expérimentales, à peine émergentes à cette époque. À la différence des articles d’Allais (1953) et Ellsberg (1961), celui d’Edwards n’est pas analysé de façon systématique dans les études sur la théorie de la décision³, et cela est compréhensible, car l’auteur ne va pas proposer ici un modèle⁴, mais plutôt suggérer ce qui se rapproche d’un programme de recherches.

Comme le remarque d’emblée l’auteur : « Economists and a few psychologists have produced a large body of theory and a few experiments »⁵. Ce qui va nous intéresser ici ce n’est pas l’article dans son intégralité, qui couvre aussi le domaine de la théorie de la décision en contexte de certitude⁶, mais la partie consacrée à la critique de l’idée de maximisation, critique fondée sur le postulat de l’utilisation des probabilités par le sujet⁷. D’autre part, les idées qu’Edwards développe dans cet article au sujet des probabilités constituent – avec les articles de 1961 et 1962⁸ – constituent un maillon important dans l’histoire de la genèse de la Théorie de la Perspective de 1979. On peut dire que d’une certaine façon, que l’article d’Edwards constitue « le complémentaire » de celui d’Allais⁹, tant pour les thèmes abordés

¹ Les apports de cet auteur en théorie de la décision sont immenses, comme en témoigne le recueil d’articles publié par Weiss/Weiss en 2009, sous le titre : *A science of decision making*.

² *Psychological Bulletin*, 51, 300-417. Cité ici : Edwards (1954).

³ Par exemple on ne trouve pas trace de cet article dans Peterson (2009), Binmore (2009), Manktelow (2012).

⁴ Ce qu’il fera par contre dans ses articles de 1955, 1963 et surtout 1968.

⁵ Edwards (1954), p 380. Il y a donc une disproportion entre la production théorique et les tests expérimentaux. Les premiers à avoir testé le modèle N/M d’un point de vue empirique sont Mosteller/Nogee (1951). On pourrait dire qu’il allait s’agir pour Edwards de faire sortir la TUA de son bureau afin de la confronter aux faits.

⁶ Pour une bonne analyse de ce concept voir Luce/Raiffa (1957), p 15-17 ; et Saint Sernin (1973), p 61-66.

⁷ Edwards propose aussi une discussion du concept de transitivité aux pages 403-405. Transitivité qui n’avait pas été analysée par Allais. Nous reviendrons sur ce concept lorsque nous étudierons les travaux de Lichtenstein et Slovic. Nous pouvons cependant dire que l’article d’Edwards anticipe les travaux de Lichtenstein/Slovic sur l’ordonnancement et la transitivité.

⁸ Il S’agit des articles : « Behavioral Decision Theory », *Annual Review of Psychology*, 12, 473-498 ; et, « Subjective Probability Inferred from Decisions », *Psychological Review*, 69, 109-135. Cet article contient des anticipations très nettes et très claires de la Théorie de la Perspective de 1979.

⁹ Dont Edwards va se désolidariser, de la même façon qu’il va se désolidariser de la pensée économique.

que pour le point de vue adopté : celui de la psychologie¹. Cet article est aussi un véritable programme de recherches empiriques dans le domaine de la théorie de la décision².

L'article débute par une critique de la façon dont l'économie – et par là même la TUA – définit l'homme et sa rationalité. Il s'agit ici surtout d'une critique de l'aspect prédictif de la théorie économique, selon cette dernière, homo oeconomicus a les propriétés suivantes : « (a) He is completely informed. (b) He is infinitely sensitive. (c) He is rational »³.

La clause (a) implique que homo oeconomicus est omniscient en ce qui concerne la totalité de ses actes potentiels et de leurs résultats⁴. D'autre part, comme le remarque Edwards dans son article de 1961 : « most of the decisions people make are not economic »⁵. Il va donc s'agir de faire sortir la théorie de la décision du domaine de l'économie. Ce qui est ici en question c'est le fait suivant : *il ne s'agit pas de nier qu'il existe des rapports et des analogies entre les décisions économiques et les autres types de décisions, au contraire, il y a des analogies structurelles fortes. Ce qui est en jeu ici c'est le fait que la théorie économique ne peut servir à expliquer tous les types de décisions, autrement dit la théorie économique sous détermine la réalité décisionnelle, en particulier sa strate psychologique.*

L'élément (b) signifie que Homo oeconomicus peut ordonner de façon complète et indéfiniment précise⁶ toutes les options qui se présentent à lui. Mais pour l'auteur, c'est l'élément (c) – présupposant (a) et (b) – qui est le plus important, cet élément qui a pour centre le concept de rationalité signifie deux choses : « He [homo oeconomicus] can weakly orders the states into which he can get, and he makes so as to maximize something »⁷. La première des deux caractéristiques nous renvoie aux axiomes d'ordonnancement et de transitivité⁸ sur lesquels l'auteur va s'interroger, et la deuxième à l'hypothèse de la maximisation de l'utilité attendue qu'Edwards va considérer de façon critique. Les caractéristiques d'homo oeconomicus vont être étudiées dans le domaine du choix en contexte de certitude et dans

¹ Il s'agit d'une critique de la théorie de la décision en contexte économique, dont le but est justement de faire sortir la théorie de ce contexte. On pourrait objecter que l'article d'Allais manifestait déjà un point de vue psychologique. Ce point de vue sera jugé insuffisant par Edwards.

² Programme qui inspirera aussi bien Slovic et Lichtenstein (1968) que Tversky (1969). Voir Lichtenstein/Slovic (2006) p 23. Il s'agit d'un programme qu'Edwards lui-même fera beaucoup avancer grâce à ses propres travaux.

³ Edwards (1954), p 381.

⁴ Cette transparence du sujet à lui-même ainsi que du cours des choses et des événements est totalement irréaliste. Elle est pourtant postulée par toutes les théories de l'utilité que nous avons analysées dans notre première partie.

⁵ Edwards (1961), p 33.

⁶ Cette propriété théorique de l'agent est due à la condition archimédienne.

⁷ Edwards (1954), p 381.

⁸ Voir nos analyses de ces clauses dans la partie sur N/M.

celui de la décision en contexte de risque. Nous allons nous intéresser seulement à la deuxième partie de l'article¹.

C'est par une distinction entre le contexte de risque et le contexte d'incertitude qu'Edwards débute son analyse. Rappelons que dans le premier cas les probabilités sont données de façon objective et pas dans le second, où elles doivent être construites par le sujet². Mais c'est surtout l'idée de maximisation³ qui va retenir l'attention de l'auteur pour le moment puisqu'il revient sur une question que nous avons déjà croisée dans notre partie sur N/M : si les sujets sont réellement des maximiseurs comment expliquer alors l'achat simultané de polices d'assurances et de tickets de loteries⁴? Ce fait, tiré de l'expérience quotidienne vient jeter un doute sur la théorie de N/M, théorie qui dans l'esprit d'Edwards possède tout de même un aspect pratique intéressant, comme l'avaient confirmé les expériences de Mosteller/Nogee en 1951⁵. Par exemple lorsqu'il s'agit dans un contexte très déterminé de calculer l'utilité cardinale d'une perspective⁶.

Ce que met en avant le problème mentionné plus haut c'est le fait que les sujets ne maximisent peut-être pas, ce qui équivaldrait à réfuter le caractère descriptif et normatif de la TUA. Il faut garder à l'esprit que le processus de maximisation⁷ suppose de la part des sujets une habileté au calcul, même élémentaire : maximiser = calculer. Et il s'agit dans la TUA de maximiser à partir de la valeur des gains. Or, comme le montrera la Théorie de la Perspective, dans la lignée de Markovitz (1952), il se pourrait fort bien que la fonction d'utilité du sujet soit

¹ Elle couvre les pages 391 à 405. La troisième partie est aussi intéressante, car elle est une analyse du livre de N/M (1947), avec des remarques intéressantes sur le minimax.

² C'est une distinction qui a été thématisée par Knight dans son livre de 1921, et qu'on trouve aussi dans le livre de Keynes de la même année.

³ Edwards reviendra sur ce thème dans son article de 1961, en connectant le problème de la maximisation avec l'aspect dynamique des processus décisionnels. Le problème capital du choix dynamique n'est pas abordé dans l'article de 1954.

⁴ À ce sujet voir Friedman/Savage (1948). Il faudra attendre la Théorie de la Perspective pour que le problème soit résolu en faisant intervenir le concept d'aversion au risque, voir Kahneman/Tversky (1979).

⁵ Voir, Mosteller/Nogee (1951). Il s'agit des premiers tests expérimentaux de la TUA. Comme le remarque Gilboa 2009, p 156, les auteurs avaient eux aussi eu l'intuition d'un coefficient de pondération des probabilités.

⁶ L'exemple que prend Edwards (1954), p 392 est intéressant : imaginons un sujet indifférent entre 7\$ sûrs et ½ de 10\$, cela implique que ces deux sommes ont pour lui la même utilité. Maintenant si nous définissons 0\$ comme 0 *utiles* (unité standard de mesure de l'utilité) et 10\$ comme 10 *utiles*, alors nous pouvons calculer l'utilité attendue que représente pour le sujet 7\$: $U(7\$) = .5U(10\$) + .5U(0\$) = .5(10) + .5(0) = 5$. 7\$ sont donc pour lui équivalents à 5 utiles. L'exemple a ceci d'intéressant qu'il permet de montrer comment la théorie fonctionne in situ sur des cas élémentaires. On trouve d'excellents exemples, parfois très élaborés dans Binmore (2009), p 47-50 ; et Peterson (2009), p 104-106.

⁷ Au cours de notre travail va émerger une tension entre l'aspect computationnel et l'aspect évaluateur à l'œuvre dans les processus décisionnels. La Théorie de la Perspective fera pencher la balance du côté de l'aspect évaluatif, contrairement à la TUA, voir Kahneman/Tversky (1979), p 28-32. Cependant on ne peut nier qu'il y ait aussi un aspect computationnel dans cette théorie, du fait qu'elle reste très proche de la TUA du point de vue formel.

basée sur son avoir et son état de fortune présent, contrairement à ce que postule la TUA, c'est-à-dire maximiser en fonction des gains probables¹.

Edwards va maintenant passer à la critique proprement dite de l'idée de maximisation. Critique qui débute par une brève remarque à propos de l'article d'Allais (1953)², jugé insuffisant du point de vue de ses apports pour la psychologie³. Pourquoi Edwards juge-t-il l'analyse d'Allais insuffisante⁴ ? Ce dont doute l'auteur, c'est que le modèle de l'économiste puisse rendre compte des choix engageant des perspectives jugées plus « normales », c'est-à-dire plus proches de celles dont l'existence se compose réellement. Au moment où Edwards écrit ce sont les expériences de Mosteller et Nogee qui avaient donné les meilleurs résultats en faveur de la TUA pour des perspectives raisonnables⁵.

Des problèmes plus sérieux allaient émerger à partir d'un autre élément de la TUA : les probabilités. En effet dans une série d'articles⁶ Edwards avait montré de façon expérimentale que les sujets sont sensibles d'une part au risque⁷ que comporte l'alternative et d'autre part à certains types de probabilités : « Subjects strongly preferred low probabilities of losing large amounts of money to high probabilities of losing small amounts of money »⁸.

¹ C'est un problème très important qui doit être relié à des considérations plus générales. En faisant l'hypothèse raisonnable que l'existence se constitue et se maintient autour de certaines habitudes, que l'on tend à acquérir ou à préserver, nous pouvons penser que nos décisions sont fondées sur l'état présent de nos systèmes de références général et décisionnel, c'est-à-dire qu'il y a une variation autour d'un noyau d'habitudes.

² Voir Edwards (1954), p 394.

³ Pourtant Allais se réclamait lui aussi de la psychologie ! Faut-il voir ici une tension entre les tendances behavioristes de l'économiste et celles plus tournées vers la psychologie cognitive émergente défendues par Edwards ? Quoi qu'il en soit, nous trouvons dans l'article des éléments qu'on retrouver chez Morgenstern (1979), notamment en ce qui concerne l'irréalisme des alternatives.

⁴ C'est probablement ces insuffisances qui ont permis à Machina (1987) de dire « The Allais Paradox was originally dismissed as an isolated example », p 129. Ce qui est en jeu ici c'est la possibilité de tester le phénomène dans la réalité, ce qu'interdisent les alternatives présentées dans la version originale. Il reviendra à Kahneman et Tversky dans leur article de 1979 de construire des alternatives basées sur le paradoxe parfaitement réalistes, voir Kahneman/Tversky (1979), p 20.

⁵ Voir, Mosteller/Nogee (1951). Il s'agissait de tester sur deux groupes de sujets venant d'horizons totalement différents s'ils étaient plus sensibles à la *valeur* attendue, ou à l'*utilité* attendue. Il s'avéra que bien qu'ayant une utilité marginale décroissante différente, les deux groupes de sujets maximisaient l'utilité attendue.

⁶ Notamment Edwards (1953).

⁷ L'auteur rejoint donc ici Allais (1953), il s'en séparera de nouveau à propos du problème de l'importance de la variance dans les options.

⁸ Edwards (1954), p 396. Ce sont donc les probabilités plus que les sommes engagées qui retiennent l'attention des sujets, la considération de la perte probable vient donc en deuxième position. Ce fait est peu compatible avec le principe de maximisation puisqu'ici les deux composantes (gains et probabilités) sont évaluées séparément. Il s'agit ici du noyau dur des expériences de Slovic et Lichtenstein (1968) et (1971) qui montreront que les deux composantes sont évaluées différemment en fonction de la tâche demandée aux sujets (jouer/évaluer monétairement), c'est à cet effet qu'ils vont d'ailleurs construire deux types de paris (« P bet » et « \$ bet »). Ce phénomène sera aussi présent dans la Théorie de la Perspective de 1979, où les probabilités subiront de surcroît une distorsion sensible à cause du coefficient de pondération. Coefficient dont Edwards avait eu l'idée bien avant Kahneman et Tversky.

Ce fait pose un problème du point de vue de la TUA qui postule que les sujets maximisent et aussi que les préférences doivent être indépendantes des probabilités. Edwards avait mis en avant un autre fait très important : le renversement des préférences qui sera repris et analysé dans toute son extension à partir des travaux de Lichtenstein et Slovic¹ ainsi que ceux de Kahneman et Tversky². Ces inconsistances des sujets par rapport aux probabilités mettent donc en difficulté le modèle N/M et conduisent Edwards à approfondir sa réflexion sur ces problèmes.

Ce qui nous intéresse ici c'est la façon dont les analyses d'Edwards vont anticiper celles de la Théorie de la Perspective, et ce dans différentes dimensions. Tout d'abord, les expériences menées par Preston et Baratta en 1948 avaient démontré que les sujets surestiment les probabilités faibles et sous-estiment les fortes probabilités³. Edwards entend ici discuter des probabilités telles que le sujet se les représente en transformant leur échelle objective. À ce propos, plusieurs questions vont émerger, et tout d'abord celle de la limitation : les probabilités objectives sont limitées par 0 et 1, doit-il en être de même pour l'échelle subjective⁴ ? D'autre part : « If the objective probability of event A is P, and that of a not occurring is Q, then $P + Q = 1$. Should this rule hold for subjective probability ? »⁵. Il s'agit ici de l'important problème de l'additivité dénombrable des événements, additivité à 1 qui constitue un axiome du calcul des probabilités⁶. Il semblerait donc que l'échelle subjective fasse subir aux probabilités objectives un certain type de distorsion, et la question que se pose l'auteur est la suivante : comment peut-on rendre compte de cette distorsion en maintenant une échelle de probabilités objectives comme échelle de référence ? Il s'agit en fait de se demander : comment remplacer le concept d'échelle subjective en formalisant un coefficient qui soit une fonction de l'échelle objective ? La réponse est importante :

¹ Voir, Slovic et Lichtenstein (1968), (1971) et (1973). Pour un point de vue plus général sur le problème on peut se reporter à leur article de 1983 ainsi qu'à la préface du volume *The construction of preference* parut en 2006.

² Voir, Kahneman/Tversky (1979).

³ Les auteurs avaient eux aussi eu l'idée d'un coefficient de pondération attaché aux probabilités, voir Gilboa (2009), p 156. Le phénomène de surestimation sera formalisé par la TP grâce aux concepts de sur-additivité et de sous-additivité, voir Kahneman/Tversky (1979), p 34-38.

⁴ Il se pourrait fort bien que cette échelle n'entretienne avec les probabilités objectives un simple rapport d'*homonymie*, et que ce ne soit tout simplement pas des probabilités que nous utilisons, mais une autre sorte d'évaluation qui resterait à modéliser, et qui se détacherait à la fois des probabilités objectives telles que posées par N/M et de la distorsion subjective des probabilités proposée par Allais, Edwards, et Kahneman/Tversky.

⁵ Edwards (1954), p 397. Voir aussi l'article de 1962 pour une discussion plus détaillée des problèmes liés à l'utilisation des probabilités.

⁶ Voir Kolmogorov (1933) ; Hacking (2004), p 67-70. Des modèles non additifs ont été construits par Schmeidler (1989).

« One way of avoiding these difficulties is to stop thinking about a scale of subjective probabilities, instead, to think of a *weight function* applied to the scale of objective probability which weights these objective probabilities according their ability to control behavior »¹.

Nous nous trouvons ici face à une idée centrale de la Théorie de la Perspective de 1979² : il faut analyser non pas comment les probabilités subjectives sont construites par le sujet – ce que fera Edwards en 1955 –, mais comment ce dernier distord et transforme les probabilités objectives, ou, pour être plus proches des concepts que nous allons mettre en place : comme le sujet assimile-t-il une information probabiliste de type objectif³ à son propre système de référence décisionnel comme faisant partie du contexte. D'autre part, à partir du moment où ce sont des coefficients de pondération qui sont pris en compte, les inconsistances avec la limitation des probabilités ou leur additivité tombent : les coefficients en questions peuvent par exemple être sous ou sur-additifs⁴.

Bon nombre d'expériences, qu'Edwards recense ici, avaient été menées pour tester empiriquement la façon dont les sujets se comportent vis-à-vis des probabilités, et par là même les pondèrent⁵. Nous ne rentrerons pas dans le détail de la revue de littérature opérée par Edwards, mais nous citerons quelques résultats importants pour la suite de notre recherche. Notamment celui-ci : il pourrait être question que les sujets produisent en fait un double coefficient de pondération ; les probabilités d'une perspective avec une utilité attendue positive ne seraient pas évaluées de la même manière que les probabilités d'une perspective ayant une utilité attendue négative, même si elles sont identiques : « The results (...) showed a very great difference between the probability preference pattern for positive and negative expected value bets »⁶.

Il s'agit ici pour nous d'un fait important, tout d'abord parce qu'il anticipe la Théorie Cumulative de la Perspective que construiront Kahneman et Tversky en 1992, théorie qui aura

¹ Edwards (1954), p 398. C'est nous qui soulignons.

² Kahneman, dans la préface du volume *Choices, Values, and Frames* ne manquera pas de reconnaître la dette qu'il a, avec Tversky, envers les travaux d'Edwards sur les problèmes de pondération, voir la préface du volume cité, p XI.

³ Comme nous le verrons avec Ellsberg (1961), les choses doivent être en fait un peu plus nuancées.

⁴ Comme c'est le cas du coefficient de pondération en Théorie de la Perspective, voir Kahneman/Tversky (1979), p 34-38.

⁵ L'auteur recense ici les expériences de Coombs sur la façon dont les sujets construisent différents types d'échelles psychologiques, voir Coombs/Milholland (1954).

⁶ Edwards (1954), p 400.

en partie pour objectif de formaliser le dédoublement des coefficients de pondération¹. Mais, d'autre part, ce fait nous amènera à nous interroger sur la possibilité d'élargir le phénomène de pondération d'autres facteurs du contexte décisionnel que les probabilités².

Autre point qui émerge, notamment à partir des travaux d'Allais (1953), et Georgescu/Roegen (1953), c'est le problème de la variance autour de la moyenne des distributions de probabilité que nous avons rencontré précédemment. Allais avait fait de cet élément (IV) un point central de sa propre théorie, mais comme nous l'avons vu, la somme et les probabilités engagées sont trop extrêmes, ce qui permet à Edwards de ne pas retenir l'approche d'Allais :

« It is enough if the theory of choice can predict choices involving familiar amounts of money and familiar probabilities differences – choices such as those which people are accustomed to making »³.

De l'avis d'Edwards, la méthode psychophysique – basée sur le concept de seuil psychologique – utilisée par Allais ne permet pas de construire une échelle d'utilité subjective satisfaisante. Il existe des éléments que la théorie d'Allais ne prend pas en compte comme le degré potentiel de surprise face à la réalisation d'une option⁴. Ce qui ressort de l'enquête critique d'Edwards, c'est toute une série de questions concernant la théorie de la décision. Pour l'auteur, les échelles d'utilité subjectives doivent pouvoir rendre compte des préférences des sujets dans le plus grand nombre de situations réelles et réalistes possibles. Tout le problème consiste à élaborer les outils pour modéliser de telles échelles.

Ce qui s'ébauche dans le paragraphe suivant, c'est la définition de ce que nous analyserons dans notre troisième chapitre comme le système de référence décisionnel. À partir

¹ Voir, Kahneman/Tversky (1992), p 48. La Théorie Cumulative de la Perspective est la généralisation « naturelle » de la Théorie de la Perspective au contexte d'incertitude et à un nombre indéfini de perspectives. Pour le caractère « naturel » de cette généralisation, voir la préface de Kahneman de *Choices, Values, and Frames*, p X. Cette extension de la théorie originale leur a en effet pris 13 ans...

² Comme nous le notions plus haut, nous pensons ici en particulier au temps, et Edwards avait lui aussi conscience de la limitation de la TUA du fait que c'est une théorie statique, voir Edwards (1961), p 34-35. Nous montrerons dans notre troisième chapitre que la pondération du temps est un élément central de toute décision.

³ Edwards (1954), p 401. Comme le remarque l'auteur juste après : « It may be necessary for economic theory that the theory of choice be universal and exceptionless, but experimental psychologists need not be so ambitious ».

⁴ En effet, cet élément n'apparaît chez Allais, ni dans les éléments fondamentaux qu'il présente, ni dans les éléments secondaires. À cette époque, seule la théorie de Shackle prenait en compte ce phénomène, tout en se passant de l'idée de maximisation.

du moment où nous disposons d'un outil comme les coefficients de pondération afin de formaliser les échelles subjectives qui mesurent l'utilité et les probabilités, nous pouvons nous demander :

« To what extent can these subjective scales be changed by learning ? To what degree people differ, and can these differences be correlated with environmental, historical or personal differences ? »¹.

Il nous faudra répondre à de telles questions pour arriver à définir de façon la plus claire possible ce que nous entendons par système de référence décisionnel. Dans le cadre de notre théorie, le SRD est envisagé comme un sous-système du SRG qui contient les éléments cognitifs susceptibles d'entrer en jeu dans les processus décisionnels². C'est précisément ce qu'a en vue Edwards lorsqu'il parle de « subjective scale ». Pour lui, l'étude des échelles d'évaluation subjectives de l'utilité et des probabilités passe par un renforcement et une élaboration plus poussée des protocoles expérimentaux, ainsi que par une multiplication des expériences. À l'époque où il écrit, les expériences en théorie de la décision étaient assez rares et souvent imparfaitement contrôlées, dans certains cas elles étaient même menées au hasard des rencontres (travail, famille, amis, etc.), donc leurs résultats étaient peu fiables :

« The results of such pilot experiments too often are picked up and written into the literature without adequate warning about conditions under which they were performed and the consequent limitations on the significance of the results. »³

Il y avait de fait une disproportion entre les théories construites et les expériences menées pour les confirmer ou les infirmer. L'article d'Edwards constitue donc une sorte de programme à la fois théorique et expérimental, programme qui sera suivi par Tversky, Lichtenstein et Slovic qui construiront leurs protocoles expérimentaux de façon nettement plus scientifique. L'article a pour but de montrer quels sont les points qui doivent être élaborés, par exemple la transitivité, dont les expériences montraient à cette époque qu'elle fonctionnait,

¹ Edwards (1954), p 403.

² Voir notre schéma, p 219.

³ Ibid.

mais pas de façon systématique, et, comme le remarque Edwards : « Utility maximisation will not work except with a transitive preference field »¹. C'est au fond l'idée de maximisation qui fait problème, car elle présuppose trop de la part des sujets, hormis dans quelques situations très particulières où elle peut s'appliquer assez naturellement. La question qu'il convient de se poser avec l'auteur est la suivante : les sujets sont-ils réellement des maximisateurs ? Mais nous allons voir que pour tenter de répondre à cette question centrale il faut passer par l'analyse d'un concept très important pour toute théorie et tout type de décision : le concept de contexte.

2) Le paradoxe d'Ellsberg et le problème du contexte décisionnel.

C'est à partir de l'analyse de l'article de Daniel Ellsberg : « Risk, ambiguity, and the Savage axioms »² que nous allons discuter, affiner et commencer de redéfinir le concept de contexte décisionnel³. Dans notre troisième chapitre nous remplacerons ce concept par deux éléments : l'environnement décisionnel et la catégorisation. D'autre part, l'analyse du paradoxe⁴ élaboré par cet auteur va nous permettre de poursuivre et de préciser la critique de la TUA, et ce de nouveau par le truchement de la contestation des prétentions normatives et prédictives du PCS, c'est-à-dire de la propriété d'indépendance. Le travail d'Ellsberg est par ailleurs lui aussi un maillon de la chaîne, qui d'Allais, nous conduit par étapes vers la Théorie de la Perspective.

Cette nouvelle réfutation « empirique » va nous permettre de mettre en lumière un concept très important dans le cadre de nos recherches : l'*ambiguïté*⁵. Ambiguïté dont on peut penser qu'elle environne bon nombre de nos décisions, et ce du fait que l'information

¹ Edwards (1954), p 405.

² *Quarterly Journal of Economics*, 75 (1961), p 643-669. Nous utilisons la version publiée dans Gärdenfors/Sahlin (1988) aux pages 245-269. Nous citons Ellsberg (1961).

³ Il s'agit d'un problème central dans le cadre de toute théorie de la décision, bayésienne ou non. Notre propos sera de redéfinir le concept de contexte d'une façon qui puisse englober plus d'informations que les probabilités.

⁴ Il s'agit par contre ici initialement d'une expérience de pensée, mais qui a donné lieu à d'importantes investigations empiriques in situ, voir Halevy (2007), Etner/Jeleva/Tallon (2010), Binmore/Stewart/Voorhoeve (2012), Colman (2011).

⁵ Il faut garder à l'esprit que la critique d'Ellsberg se situe de façon stricte dans le paradigme bayésien, et que l'ambiguïté sera définie par lui dans ce contexte, voir Epstein/le Breton (1992). Etner/Jeleva/Tallon (2010), p 34, remarquent que l'ambiguïté peut parfois être utilisé comme synonyme de l'ignorance en contexte décisionnel. Il nous semble que tel n'est pas le cas dans l'article que nous étudions, puisque dans ce cas-là il suffirait de traiter les deux concepts de façon équivalente et nous nous retrouverions avec une théorie très proche de celle de Savage, or précisément, du fait que l'article d'Ellsberg est une critique de Savage, ambiguïté et ignorance ne peuvent être tenues pour synonymes. Ellsberg montrera d'ailleurs aussi que l'ambiguïté ne peut être réduite à l'incertitude par l'application du Principe de Raison Insuffisante (PRI).

disponible n'est ni toujours optimale ni toujours utilisée à son maximum. Il s'agira pour nous ici, comme ce fut le cas avec Edwards, non pas d'analyser l'article in extenso dans toutes ses ramifications et implications¹, mais d'exploiter certains thèmes et certaines pistes de réflexion proposés par l'auteur.

Comme nous le remarquons en note, la plupart des études² consacrées à l'article de 1961 se concentrent sur le paradoxe et sur la mise à jour de l'aversion pour l'ambiguïté, en ne mentionnant que très peu les analyses préliminaires que l'auteur propose sur le concept de contexte. Contexte défini de façon classique par le rapport du sujet aux probabilités. En effet, comme le remarque Ellsberg, même en contexte d'incertitude : « The feeling has persisted that people tend to *behave* 'as though' they assigned numerical probabilities or 'degrees of belief', to events impinging their actions »³. Il s'agit donc de faire l'hypothèse que la définition du contexte est liée de façon intime aux distributions de probabilités⁴. Rappelons que la distinction « classique » entre les contextes décisionnels avait été élaborée par F. Knight en 1921⁵ et qu'elle est la suivante⁶ :

- Contexte de risque où l'événement arrive avec des probabilités objectives, comme c'est le cas dans les axiomatiques de N/M (1947), Marschak (1950), Samuelson (1952). On peut donc dans ce cas modéliser les options sous la forme de tickets de loterie : $(x, p ; y, 1 - p)$.
- Contexte d'incertitude ou d'ignorance dans lequel le sujet doit lui-même construire son évaluation probabiliste, par exemple à partir de ses croyances, comme c'est le cas par exemple chez Ramsey (1926), Savage (1954) et Jeffrey (1965). Dans ce cas-là on préférera le concept de pari à celui de loterie.

Comme nous l'avons mentionné, l'utilisation des probabilités – quel que soit leur

¹ Travail qui n'a pas encore été mené de façon exhaustive à notre connaissance, puisque les articles ou les ouvrages consacrés à ce problème se limitent le plus souvent à la présentation et à l'explication du paradoxe.

² Par exemple Baron (1988) ; Willinger (1990) ; Epstein/le Breton (1992) ; Picavet (1996) ; Halevy (2009) ; Peterson (2009) ; Etner/Jeleva/Tallon (2010) ; Colman (2011) ; Binmore et Al. (2012).

³ Ellsberg (1961), p 245.

⁴ Voir Luce/Raiffa (1957), p 13-23 ; Peterson (2009), p 40-86 ; Binmore (2009), p 35.

⁵ Dans son livre *Risks, uncertainty and profit*. C'est-à-dire dans le contexte de l'économie. D'autre part, on trouve des idées proches de Knight dans Keynes (1921), voir Colman (2011), p 2.

⁶ Nous éliminons ici le contexte de certitude comme non pertinent pour nos travaux, voir Saint-Sernin (1973), p 61-66.

interprétation – pose problème, problème qu’Ellsberg mentionne d’ailleurs, en critiquant par exemple l’application de la loi des grands nombres au contexte décisionnel¹ : comment et quelles probabilités appliquer dans le cadre d’un événement strictement unique ? Car une théorie comme celle de N/M n’a de sens que si l’utilité des perspectives est mise en relation avec la loi des grands nombres. C’est-à-dire que la valeur d’une perspective ou d’un ticket de loterie ne s’approchera de sa valeur moyenne que sur le long terme. La théorie de Ramsey/Savage est exposée à une autre forme de problème : si l’on accepte la révision des probabilités subjectives via le théorème de Bayes que se passe-t-il dans le cadre d’événements uniques dont la probabilité et par là même la croyance ne peut être sujette à révision ? La mise à jour bayésienne suppose elle aussi, non pas une répétitivité de l’événement, mais au moins une évolution temporelle de l’information.

D’une façon plus centrale pour notre propos, et si l’on veut faire *sortir* le concept de contexte de l’économie et lui donner une densité psychologique et philosophique, il va falloir le redéfinir. Aussi posons-nous cette double question préalable² : le contexte décisionnel peut-il être réduit à l’état de connaissance des sujets quant aux probabilités ? Les sujets utilisent-ils vraiment les probabilités³ ?

L’article d’Ellsberg va mettre en lumière que même dans un contexte strictement bayésien on ne peut répondre à la première question par l’affirmative. Notamment en mettant l’accent sur le fait que certaines incertitudes ne sont pas mesurables et ne peuvent être transformées en risque⁴ ainsi que pourrait le faire penser le Principe de Raison Insuffisante (PRI)⁵. Comme se le demande l’auteur :

¹ Ellsberg (1961), p 246 ; sur ce point voir aussi Gigerenzer (1991).

² Ces questions recevront réponse dans le troisième chapitre.

³ Questions qui en appellent une troisième : existe-t-il réellement une information de type probabiliste ou statistique disponible et utilisable par le sujet dans toute situation décisionnelle ? Et même dans le cas où la réponse soit affirmative, les sujets sont-ils capables de traiter et d’utiliser cette information ? C’est ce que mettront en doute Kahneman et Tversky avec leur théorie des heuristiques et des biais de 1974, sur laquelle nous allons revenir sous peu. Cette hypothèse – qui tend à sous-entendre que les sujets sont « défaillants » du point de vue du rôle normatif des statistiques et des probabilités – sera critiquée par Cohen (1981) ; Cosmides/Tooby (1996) ; et Gigerenzer (1991). Les hypothèses des deux derniers articles tenteront de faire sortir les sujets de la « misère cognitive » dans laquelle la théorie des heuristiques et des biais les avait plongés, il s’agira entre autres d’intégrer le concept d’adaptativité et d’environnement dans le processus décisionnel. Nous reviendrons sur ces hypothèses au début de notre troisième chapitre. Pour un bon état des lieux sur ce débat, voir Andler (2004), p 368-400.

⁴ Ellsberg (1961), p 248 et suivantes.

⁵ Principe qui avait été explicité par Bernoulli et Laplace, et dont le sens est le suivant : si il n’y a pas de raison contraignante on doit distribuer de façon égale les probabilités entre les événements. Soit l’événement : demain il pleuvra (A) et son complément : demain il ne pleuvra pas ($\neg A$) ; sans raison suffisante qui me dicte le contraire le PRI stipule que $p(A) = \frac{1}{2}$ et $p(\neg A) = \frac{1}{2}$. Voir Gigerenzer (1988), p 6-11 ; Daston (1989), p 29-30 ; Hacking (2004), p 156 ; et Peterson (2009), p 53, qui explique comment tout contexte d’ignorance en théorie de la décision peut être réduit à un contexte de risque par l’application du PRI.

« In considering only deliberate decisions, then, does this leave any room at all for ‘unmeasurable uncertainty’ : for uncertainties *not* reducible to ‘risks’, to quantitative or qualitative probabilities ? »¹

L’aspect normatif des axiomatiques N/M et Savage repose de façon explicite sur la définition d’un contexte, lui-même relié à une information probabiliste. Qu’en est-il de ces théories lorsque nous nous trouvons confrontés à des contextes décisionnels qui n’appartiennent ni à l’incertitude ni au risque ? C’est ce que va mettre en avant le paradoxe. Mais avant d’en venir à lui, approfondissons la notion de contexte. Il apparaîtra en fait au fur et à mesure que nous avancerons que dans le contexte décisionnel interviennent d’autres facteurs que les probabilités. Facteurs dont il faudra identifier les principaux éléments. D’autre part, nous aurons l’occasion de montrer que les probabilités n’interviennent de façon pertinente que dans certains types de décisions particulières où les pratiques et l’exercice répété ont justement permis d’établir certaines informations statistiques fiables².

Il nous faudra donc, arrivés à un certain point de notre réflexion, *déconnecter le concept de contexte de celui de probabilité*, et le rapprocher d’autres facteurs comme le temps par exemple ou le concept de délai, et de bien d’autres éléments importants comme l’environnement médiat ou immédiat³. Nous montrerons que le concept de contexte tel qu’utilisé en théorie de la décision est trop restrictif, et qu’il faut le faire évoluer vers le concept *d’environnement décisionnel*⁴ qui laisse plus de place à la complexité, à la plasticité, à l’adaptativité, et à l’influence déterminante du système de référence subjectif dans la prise de décision.

Ellsberg, après avoir présenté les quatre premiers axiomes de Savage, se prête à deux expériences de pensée⁵ qui ont pour effet de remettre en cause P1, P2 et le PCS, mais aussi P4 indirectement. La démarche de l’auteur est originale, même en acceptant l’aspect normatif de

¹ Ellsberg (1961), p 247.

² C’est le cas par exemple dans le domaine médical et dans les domaines qui se rapproche structurellement de lui. C’est-à-dire les pratiques qui supposent un exercice constant qui a lui-même évolué au cours des âges, tel le sport notamment. Au sujet du domaine médical et des décisions qui y sont afférentes, on peut consulter les exemples de Baird (1992), p 50-75 ; voir aussi Abecassis (1999).

³ Par ailleurs si les probabilités doivent être maintenues, il faudra se questionner sur leur structure qui n’obéira pas de façon stricte aux axiomes classiques. On pourrait parler de probabilités métaphoriques se situant quelque part entre la computation et la représentation, des probabilités que pourrait saisir par exemple la notion usuelle de « chance de ».

⁴ C’est dans ce sens que vont les travaux de Simon (1955) et Gigerenzer/Brighton (2009).

⁵ Pages 252 et 255

l'axiomatique¹, c'est-à-dire en acceptant du même coup la distinction classique des contextes, on peut construire des situations hypothétiques dans lesquelles les axiomes sont inapplicables pour des raisons techniques : il manque au sujet un certain type d'information² concernant les probabilités d'un événement de référence. Soit la situation suivante : « Imagine an urn known to contain 30 red and 60 black and yellow balls, the latter in unknown proportion »³. En élaborant la situation nous pouvons proposer les tableaux suivants⁴ :

Première situation :

	30 boules	60 boules	
	Rouges	Noires	Jaunes
I	100\$	0\$	0\$
II	0\$	100\$	0\$

Dans la première situation les choix sont présentés ainsi : **I** parier sur rouge et gagner 100\$ si une boule rouge est tirée, rien autrement, **II** parier sur noir et gagner 100\$ si une boule noire est tirée. Seule la proportion de boules rouges est connue, celle des boules noires et jaune est comprise quelque part entre 0 et 60. On demande aux sujets de dire s'ils préfèrent l'option **I** ou **II**.

Deuxième situation :

	30 boules	60 boules	
	Rouges	Noires	Jaunes
III	100\$	0\$	100\$
IV	0\$	100\$	100\$

¹ Ce que fait Ellsberg : « In general, as one ponders these postulates and test them introspectively in a variety of hypothetical situations, they do indeed appear plausible », p 252.

² Nous aurons bientôt l'occasion de nous rendre compte que le concept d'information est relié à celui de représentation, notamment quand nous étudierons les travaux de Kahneman/Tversky et Gigerenzer.

³ Ellsberg (1961), p 255.

⁴ Pour de bonnes analyses, voir Baron (1988), p 281-285 ; Willinger (1991), p 15-16 ; Picavet (1996), p 230-233 ; Peterson (2009), p 81-83.

Dans la deuxième situation les deux rétributions en cas de tirage d'une boule jaune sont changées de 0\$ à 100\$. On demande là aussi aux sujets quelle option ils préfèrent. La réponse la plus fréquente est que **I** est préféré à **II** et que **IV** est préféré à **III**, ce qui est une violation directe du PCS, puisque la seule chose qui change d'une situation à l'autre c'est la conséquence commune aux deux situations : 0\$ dans la première et 100\$ dans la seconde. De façon moins fréquente, mais là aussi en contradiction avec ce que prédit le PCS on a les préférences suivantes : **II** est préféré à **I** et **III** est préféré à **IV**.

Analysons de plus près ces résultats¹. Avec quels axiomes de la théorie de Savage sont-ils incompatibles ? Comme nous le notions, le PCS est immédiatement touché et de façon évidente, comme dans le paradoxe d'Allais et pour les mêmes raisons. Ce qu'il faut considérer c'est le rapport des probabilités et des croyances avec l'axiome d'ordonnancement. Comme le remarque Ellsberg : « It is impossible, on the basis of such choices, to infer even qualitative probabilities for the events in question (specifically, for events that include yellow or black, but not both »².

Rappelons-nous que chez Savage, dans l'axiome 4, les probabilités sont construites à partir des croyances du sujet en la réalisation ou non d'un événement de référence³. Or, dans le cas que nous analysons, nulle mesure de probabilité cohérente ne peut être inférée du choix des sujets, si ceux-ci reflètent leurs croyances. En effet posons que A = l'événement tirer une boule rouge, et B = l'événement tirer une boule jaune ou noire. Avec des préférences telles qu'exprimées par les sujets nous nous retrouvons dans la situation suivante : pour la première alternative, avec $>'$ désignant une relation de croyance, nous avons $A >' B$, et pour la deuxième alternative $A^c >' B^c$. Ce qui revient à dire – si les probabilités subjectives sont fondées sur les croyances comme le stipule l'axiome 4 de Savage – que :

¹ Comme nous le notions plus haut, à l'origine le paradoxe était une expérience de pensée, ce qui pourrait légèrement limiter sa portée théorique. Seulement l'expérience de pensée a donné lieu à des expériences réelles dont les résultats indiquent différentes directions théoriques. La robustesse de l'aversion pour l'ambiguïté a été mise en évidence, entre autres, par Halevy (2007) qui tenta de l'expliquer à l'aide de différents modèles théoriques, pour plus de détails voir p 508 et 511-518. De leur côté Binmore et Al. (2012) ont montré qu'avec des probabilités légèrement modifiées par rapport à l'article d'Ellsberg, les préférences des sujets reflétaient l'utilisation presque systématique du PRI plutôt que du critère d'Hurwicz. Un autre résultat important, obtenu par Colman (2011), qui s'inspire d'ailleurs de Halevy (2007), indique que les sujets font preuve d'une aversion pour le l'ambiguïté si le choix est unique ; cette aversion a tendance à s'évanouir sur le long terme, dans le cas où les alternatives sont présentées sous la forme de jeux répétés, c'est-à-dire que les sujets sont cohérents du point de vue de la consistance dynamique. Colman imagine différents types de jeux répétés dans lesquels les effets du paradoxe disparaissent. Il s'agit de montrer que le paradoxe ne contredit pas la TUA.

² Ellsberg (1961), p 256.

³ Voir Savage (1954), p 30-33.

$$P(A) > P(B)$$

et

$$[1 - P(A)] > [1 - P(B)]$$

C'est là une contradiction manifeste qui interdit la construction de toute mesure probabiliste. Nous avons donc ici à la fois une contestation de l'axiome 4 et de l'axiome d'ordonnancement puisque ce qui se produit c'est un renversement des préférences. Renversement qui viole la propriété d'asymétrie de la relation d'ordonnancement.

D'autre part, les résultats¹ de cette expérience démontrent de façon directe que le PRI n'est pas appliqué puisque dans ce cas-là il devrait y avoir indifférence entre les deux paris puisque dans les deux situations : **I** = (1/3)100\$, **II** = (1/3)100\$ et, **III** = (2/3)100\$, **IV** = (2/3)100\$. Or nous n'observons aucune indifférence, et même dans le cadre des expériences qui ont été menées réellement, avec ces probabilités, les incompatibilités avec la TUA restent robustes². Les préférences des sujets démontrent donc une aversion pour l'ambiguïté, aversion dont on peut penser qu'elle est liée à une carence dans de l'information :

« What is at issue might be called the *ambiguity* of this information, a quality depending on the amount, type, reliability and 'unanimity' of information, and giving rise to one's degree of 'confidence' in an estimate of relative likelihood »³.

L'ambiguïté émerge donc ici comme un facteur critique dans l'élaboration du choix rationnel. Et, comme le remarque l'auteur, elle est liée à la quantité et à la qualité de l'information que le sujet peut retirer du contexte décisionnel. À cet égard quelques remarques s'imposent. La première c'est que lorsque les sujets choisissent comme ils le font dans le paradoxe, leur choix ne reflète l'utilisation d'aucun critère issu de la théorie de la décision : maximiser, minimaxer, appliquer le critère d'Hurwicz, etc.⁴ Comme le remarque l'auteur : « *None of the familiar criteria for predicting or prescribing decision-making under uncertainty*

¹ Résultats qui ont été contestés par Binmore et Al. (2012), avec des probabilités légèrement modifiées.

² Voir Etner et Al. (2010), p 260.

³ Ellsberg (1961), p 258.

⁴ L'article de Binmore et Al. (2012) étudie justement les rapports entre le paradoxe et ces critères décisionnels, au bénéfice du PRI, et ce à la surprise des auteurs, voir p 20.

corresponds to this patterns of choices »¹. La question qu'il convient de se poser est celle-ci : doit-on pour autant considérer les sujets qui font de tels choix comme irrationnels² ? Probablement pas, leurs choix indiquent clairement que le concept d'information joue un rôle central. Cela n'aurait aucun sens de qualifier d'irrationnel un comportement qui reflète le fait que les sujets prennent en compte la qualité et la quantité de l'information.

Les données issues du paradoxe nous conduisent aussi à réfléchir sur le concept de contexte décisionnel et à nous demander dans quelle mesure les données probabilistes peuvent-elles suffire à le déterminer ? Car le paradoxe nous montre que les sujets ne se représentent pas l'incertitude de façon probabiliste, sans quoi ils appliqueraient le PRI, ou tout autre critère. Les représentations des sujets sont donc d'un autre ordre, et le concept de contexte lui-même demande à être repensé. Il est à noter que les liens entre le contexte décisionnel et l'information ont commencé à être établis de façon systématique lorsque Simon (1956) a connecté le concept de décision avec celui d'environnement³.

Le sujet a besoin de plus d'information que l'information probabiliste : l'environnement décisionnel est, pour chaque sujet, d'une telle complexité que les probabilités sont nécessairement insuffisantes pour le construire exhaustivement. De son côté Edwards (1962) et (1968) fera lui aussi l'hypothèse – par le biais de la formalisation de la révision des croyances – que l'information joue un rôle central⁴. Seulement ses hypothèses restent dans le cadre du bayésianisme et font donc un usage important des probabilités⁵.

Il va donc nous falloir modifier le concept de contexte décisionnel de façon à y intégrer d'une part l'information susceptible d'être retirée par le sujet de ce qui l'entoure de façon immédiate ou non, d'autre part, les croyances et les valeurs des sujets, qui sont dynamiques et qui constituent le système de référence général. Plus nous avançons vers les décisions, telles qu'elles sont prises dans la réalité, dans la quotidienneté et plus nous sondons les limites de la théorie de la décision tant au point de vue descriptif que normatif. Aussi est-ce ce que l'article d'Ellsberg met en avant : dans certaines situations, la violation des axiomes de Savage ne peut

¹ Ellsberg, *ibid.*

² Même s'ils maintiennent leurs préférences après explication. Voir Ellsberg (1961), p 269.

³ Voir Simon (1956), p 129-131.

⁴ Cette idée est aussi in nucléo dans Savage (1954) qui est d'ailleurs un des co-auteurs de l'article de 1968.

⁵ En fait l'idée que nous soutenons ici ce n'est pas que les probabilités doivent nécessairement disparaître du contexte décisionnel, mais que leur utilisation est infiniment plus parcimonieuse que le bayésianisme le laisse entendre, bayésianisme qui ne représenterait donc qu'une partie de la théorie de la décision. Pour un point de vue exhaustif sur l'utilisation des probabilités dans le bayésianisme, voir Oaksford/Chater (2007) et Chater/Oaksford (2008).

être taxée d'irrationnelle, et nous aurons l'occasion de constater que c'est en effet ce qui se passe la plupart du temps, car l'ambiguïté et le manque d'information n'accompagnent-ils pas la plupart de nos décisions – dont on pourrait dire qu'ils sont la marque de fabrique – sans que cela ne menace pourtant la cohérence de nos existences. Mais cette cohérence sera alors à chercher ailleurs que dans la conformité des décisions aux règles axiomatiques.

3) Le problème du renversement des préférences (RP)

Contrairement aux travaux que nous venons d'étudier, les recherches entreprises par Sarah Lichtenstein et Paul Slovic (L/S), dans la lignée des analyses d'Edwards, sur le renversement des préférences en 1968¹ n'avaient pas pour finalité première de contester les propriétés descriptives et normatives de la TUA², mais de montrer comment les sujets construisent leurs préférences au moment où advient la situation. L'aspect critique des travaux de L/S a émergé parallèlement à leurs recherches, et ce de façon nécessaire, car les résultats des expériences sur lesquelles ces études étaient fondées montraient que les sujets violaient de façon systématique les axiomes d'ordonnancement et de transitivité³ : « This behavior is of interest because it violates almost all theories of preference, including expected utility theory »⁴. Avant de passer à l'analyse des travaux de L/S, nous allons poser deux questions programmatiques : qu'entend-on par renversement des préférences (« preference reversal »⁵) ?

¹ Le premier article coécrit par ces deux auteurs paru cette année, il avait pour titre : « Relative Importance of Probabilities and Payoffs in Risk Taking », *Journal of Experimental Psychology Monograph*, number 3, part 2, p 1-18. Nous citerons : Slovic/Lichtenstein (1968). Cet article séminal fut suivi de deux autres articles en (1971) et (1973), dont nous dirons quelques mots.

² Slovic/Lichtenstein (1968), p 49. Nous citons l'article à partir de sa réédition abrégée dans l'ouvrage coédité par les deux auteurs en 2006 : *The Construction of Preference*.

³ La critique proprement dite arrivera avec l'article de 1971 : « Reversals of Preference Between Bids and Choices in Gambling Decision. », *Journal of Experimental Psychology*, vol. 89, p 46-55. Nous citerons L/S (1971), dans la réédition incluse dans le livre de 2006 cité plus haut aux pages 52-68.

⁴ Slovic/Lichtenstein(1983), il s'agit d'un article qui fait une mise au point historique et qui met le phénomène de RP en rapport avec la Théorie de la Perspective qui est aussi confrontée à ce phénomène et a pour but de le formaliser, voir p 600-601. À l'époque où cet article fut écrit, les travaux de Goldstein et Einhorn sur le RP n'avaient pas encore vu le jour, or ces travaux auront précisément pour but d'utiliser la phase d'édition de la Théorie de la Perspective pour tenter de résoudre le problème. Voir Goldstein/Einhorn (1987), p 240.

⁵ Le sujet en tant que tel a eu un impact très large dans les domaines de l'économie, de la psychologie cognitive et de la théorie de la décision, comme en témoigne le volume coédité par L/S en 2006 : *The construction of preference*. Le phénomène RP avait déjà été mis en évidence par Allais et Ellsberg, car leurs « paradoxes » produisent en fait un RP, voir Schmeltzer et Al. (2008). L'impact a surtout été ressenti dans le domaine de l'économie, en effet, le RP réfute l'aspect prédictif de la TUA sur laquelle repose une partie de la théorie économique. Certains économistes ont tenté, en vain, de montrer que ce phénomène n'était pas robuste en construisant des protocoles expérimentaux différents, mais les résultats effectifs produisirent l'effet inverse des résultats attendus et confirmèrent massivement la robustesse du RP. Voir à ce sujet, l'article classique de Grether/Plott (1979). Le phénomène fut par contre intégré à l'économie comportementale par R. Thaler (1980).

Quel est le lien entre préférence et décision ? Lien qu'il va falloir peut-être construire philosophiquement. La définition générale du phénomène de RP est la suivante : « we could construct pairs of gamble, A and B, such that A would be preferred under one response mode and B under another »¹. Il s'agira maintenant de déterminer quels types de paris ont été utilisés par les auteurs. Il fallait construire des paires de paris composés d'options engageant différents stimuli². Cette opération présuppose que l'attention du sujet soit donc divisée entre les différentes sources d'informations qui lui sont proposées dans les paris. Ceci aurait tendance à appuyer le fait que les décideurs ne considèrent pas les options comme des totalités, mais les recomposent subjectivement à partir de leurs éléments³. Un RP aura lieu quand un certain nombre de conditions expérimentales vont se trouver réunies :

« individuals are presented with two gambles, one featuring a high probability of winning a modest sum of money (The P bet), the other featuring low probability of winning a large amount of money (the \$ bet). The typical finding is that people often choose P bet but assign a larger monetary value to the \$ bet »⁴.

Dans le cadre des articles de 1968 et 1971⁵ – les options étant présentées par paires dans un premier temps –, on demandait aux sujets deux choses différentes : évaluer auquel des deux paris ils aimeraient jouer et celui qu'ils estimeraient avoir une valeur monétaire plus importante, c'est-à-dire le prix de vente minimum auquel ils consentiraient à vendre leur ticket. Pour les paris avec une utilité négative on demandait aux sujets combien il faudrait les payer pour jouer en fonction de l'utilité attendue du pari. Il s'agit donc de demander aux sujets d'accomplir deux tâches différentes, mais du point de vue de la théorie de l'utilité ces deux tâches auraient du mener aux mêmes préférences. Revenons sur le protocole que L/S avaient

¹ L/S (2006), p 5. Le concept clé est ici celui de « response mode », il s'agira en effet de demander aux sujets différentes tâches, par exemple jouer/évaluer financièrement.

² Le stimulus est un concept central de la théorie de la décision, et son importation du registre biologique est parfaitement justifiée, comme les remarquent les auteurs : « Gambles commonly are characterized as *multidimensional stimuli* », L/S (1973), p 41. C'est nous qui soulignons, il pourrait tout aussi bien s'agir de la définition du concept de décision.

³ Cela sera confirmé par la théorie de la perspective de Kahneman et Tversky (1979), par la découverte des effets de cadrage, et enfin par la théorie de l'expression de Goldstein et Einhorn (1987).

⁴ Slovic/Lichtenstein (1983), p 236-237. Pour une autre description très intéressante du phénomène voir Goldstein/Einhorn, p 236-237. Il s'agit de l'important article sur la théorie de l'expression, en partie basée sur la théorie de la perspective (pour le concept de codage), cette théorie vise à expliquer le RP en distinguant trois phases dans la construction des préférences : codage, évaluation et expression. Le RP advient dans la troisième phase où le sujet doit exprimer ses préférences dans une échelle subjective.

⁵ L/S, (1971), p 53.

mis en place : « For each pairs, S indicated which bet he would prefer to play. After S made his choice, he then made a bid for each of the bets, which this time were presented one at time »¹. La dernière partie de la citation appelle une remarque immédiate : dans un premier temps on présente aux sujets tous les billets simultanément. Il s'agit donc d'une *évaluation jointe*, dans un second temps, les billets sont présentés un par un, il s'agit donc d'une *évaluation séparée*. La question que se sont posés et que se posent encore aujourd'hui les chercheurs vu la résistance du phénomène, est de savoir si le mode de présentation des options n'est pas déterminant, voire même s'il n'est pas la cause du PR². La robustesse du phénomène a amené des psychologues, par exemple Tversky (1988), à le considérer comme un biais cognitif. D'un point de vue normatif, le prix le plus haut devrait correspondre à ce que l'on préfère, dans le domaine de l'économie les deux tâches sont censées révéler les mêmes préférences. La structure du phénomène est en contradiction directe avec l'axiome d'ordonnement³, mais aussi avec la propriété de transitivité⁴.

Il est important de noter que le RP était déjà à l'œuvre dans les paradoxes d'Allais et d'Ellsberg. Comme nous allons le voir sous peu, la Théorie de la Perspective vise aussi à expliquer et à formaliser certaines formes du phénomène⁵. Le RP est donc avec le problème de la pondération des probabilités et la critique du principe d'indépendance un fil conducteur qui permet de relier les travaux que nous étudions dans cette partie, et qui vont d'Allais à la Théorie de la Perspective.

Avant d'analyser d'un peu plus près l'article fondateur de 1968, interrogeons-nous sur le lien entre préférence et décision. C'est un lien qui semble aller de soi puisque nous pourrions simplement dire que nos décisions sont le reflet de nos préférences, ou plus précisément que nos décisions expriment les préférences qui les motivent. Cependant, en

¹ L/S (1971), p 53.

² Voir Hsee (1996) ; Hsee et Al. (1999) ; Willemsen/Keren (2004).

³ En particulier avec la propriété d'antisymétrie de la relation d'ordonnement. Voir notre partie sur NM ; voir aussi Fishburn (1970) ; Kreps (1988), p 7 et suivantes.

⁴ Voir Slovic/Lichtenstein (1983), p 599-600 et leur commentaire à propos des expériences de Tversky (1969) sur l'évaluation par aspect. Cependant l'explication la plus claire de la façon dont le RP viole la transitivité se trouve dans Goldstein/Einhorn (1987), p 237. Explication que nous reprenons ici : supposons qu'un sujet attribue un prix minimum $a\$$ pour un objet A, et $b\$$ pour un objet B de telle sorte que $a\$ > b\$$. Si on lui donne le choix entre A et une somme $c\$$ telle que $a\$ > c\$ > b\$$, du point de vue normatif il doit choisir A. Choisir $c\$$ plutôt que A pourrait être considéré comme vendre A pour $c\$$ qui se situe en dessous du prix minimum de vente de A. D'autre part, si le sujet a le choix entre $c\$$ et B, le point de vue normatif nous indique qu'il doit choisir $c\$$, puisque cette somme est supérieure au prix de vente minimum de B. Comme A devrait être préféré à $c\$$ et $c\$$ à B, la transitivité requiert que A soit choisi plutôt que B. Or les expériences ont montrés de façon massive que la plupart, et bien qu'évaluant $a\$ > b\$$, les sujets choisissent B plutôt que A.

⁵ Voir L/S (2006), p 26-28 ; Slovic/Lichtenstein (1983), p 600-601 ; voir aussi Kahneman/Tversky (1979), p 22.

analysant la relation de plus près on s'aperçoit qu'il s'agit en fait d'une relation complexe. En effet, il arrive aussi qu'une préférence donnée soit le reflet ou l'expression d'une décision.

Par exemple, un sujet ayant décidé de se remettre à pratiquer un entraînement sportif régulier, il préfère pratiquer le vélo d'appartement que la course à pied. Mais cette décision, à son tour est fondée sur une ou plusieurs préférences antérieures : il préfère être en santé et se maintenir par le biais du sport. La question est de savoir, si nous retraçons la chaîne causale qui va des préférences aux décisions et réciproquement jusqu'à son origine, à quoi aurons-nous affaire ? À une préférence ou à une décision ? À un ensemble de préférences ? Les liens entre ces deux concepts sont complexes et multiples. Nous pouvons dire que les décisions et les préférences s'entre-expriment les unes les autres, avec cette double réserve que nous présentons sous forme de question : toute décision est-elle l'expression d'une préférence, toute préférence est-elle le reflet d'une décision ? C'est un problème en soi qu'il conviendrait d'approfondir.

Il semble que certaines préférences assez spontanées puissent être désolidarisées de toute décision, par exemple les préférences gustatives ou esthétiques, elles peuvent cependant être modifiées par le biais de la décision. Par exemple, le sujet peut se forcer à manger un aliment qui lui déplaît, par politesse, ou parce qu'il sait qu'il a certaines vertus importantes pour sa santé. Il est par contre beaucoup plus difficile de trouver des décisions qui ne fassent fond sur aucune préférence, même dans les décisions prises en toute « impartialité » comme les décisions judiciaires, ou comme la décision sceptique de suspendre son jugement, suspension reposant sur une préférence tacite qui par là même hypothèque l'indifférence du sceptique¹.

Décision et préférence sont par contre liées avec les concepts de représentation et de croyance, mais dans le cas de la préférence le rapport à la croyance n'est pas tourné de façon immédiate vers l'action, comme c'est le cas pour la décision². Nous pourrions dire de façon peut-être provisoire qu'une décision est une préférence qui quitte le domaine de la pure subjectivité pour aller s'inscrire et se refléter dans le monde.

¹ Voir Long/Sedley (1987), p 37-59. Il est certain que c'est un problème qui devrait être plus amplement développé, et le problème du scepticisme reste d'actualité, notamment en épistémologie.

² Action qu'il s'agirait de définir de façon précise, toute action n'est pas nécessairement mouvement : l'élève qui apprend par cœur sa poésie agit. Pour une bonne analyse du registre du faire en relation avec la volonté voir Ricoeur (1950), p 253-258 ; Proust (2005), p 33-45.

Une thèse importante de L/S est que nos préférences se construisent à l'occasion de la situation : « The central idea is that in many situations we do not really know what we prefer ; we must construct our preferences as the situation arise. »¹. C'est entre autres dans le but de tester cette idée qu'ont été élaborées les expériences décrites dans l'article de 1968 : « Relative Importance of Probabilities and Payoffs in Risk Taking »².

Comme le remarquent d'emblée les auteurs : « Gambles commonly are characterized as multidimensional stimuli »³. Le concept de stimulus, tiré du domaine de la biologie est ici à considérer, car il fait référence à ce qui déclenche la construction des préférences, au stimulus répond un certain comportement que les auteurs vont analyser. Remarquons qu'ici, comme c'était déjà le cas chez Allais et Ellsberg, c'est la représentation qui est à l'œuvre. Dans tous les cas, les sujets n'ont jamais affaire aux alternatives telles qu'elles sont données, mais à leurs représentations⁴.

Dans le cadre des paris présentés lors de l'expérience de 1968 il y a quatre dimensions liées au risque à considérer qui vont jouer le rôle de stimuli. Par exemple pour des paris doubles (« duplex gambles »), présentés aux sujets lors de la première expérience, on trouvait les quatre dimensions suivantes : probabilité de gagner (Pg), montant à gagner ($\$g$), probabilité de perdre (Pp), montant à perdre ($\$p$). Il s'agira donc de construire des expériences qui permettront d'analyser comment ces quatre stimuli influencent les préférences des sujets lorsqu'on leur demande :

1 – À quelle loterie ils préféreraient jouer.

2 – Quelle est la loterie qui à leurs yeux a la plus grande valeur monétaire.

¹ L/S (2006), p 1. Nous verrons plus loin, notamment avec l'étude de la rationalité adaptative que ce fait semble confirmé. Voir Gigerenzer/Brighton (2009) in Gigerenzer (Ed.) (2011), p 19-22.

² Nous citons la version abrégée de cet article publié dans L/S (2006), aux pages 41-51. À cet article séminal il faut ajouter celui de 1971 où les protocoles expérimentaux furent encore plus élaborés, et celui de 1973 qui rapporte des expériences menées dans le casino « Fourth Queen » de Las Vegas.

³ Slovic/Lichtenstein (1968), p 41.

⁴ Voir à ce sujet les remarques de Kahneman, dans la préface de *Choices, Values, and Frames*, p XIV : « anyone who has taken a course in perception has learned to distinguish objective reality from the proximal stimulus to which the observer is exposed and to distinguish both reality and the stimulus from the mental representation that the observer eventually constructs ». C'est un texte très important sur lequel nous reviendrons lorsque nous approfondirons la problématique de la représentation. En fin de compte, tout se joue à l'intérieur de la sphère représentationnelle. Il y a un aspect assez kantien derrière cette notion de représentation.

Dans le cadre de la première expérience¹, un pari typique peut être présenté comme suit : .4 de gagner 1\$, .6 de ne rien gagner, .2 de perdre 4\$, et .8 de ne rien perdre². Nous voyons donc que les paris sont construits en fonction des quatre dimensions décrites plus haut, et à l'origine la question était la suivante : quelle est ou quelles sont les dimensions qui pèsent le plus dans la construction des préférences³ ?

L'expérience a été menée sur 213 sujets qui étaient payés, mais qui ici ne jouaient pas pour de bon, contrairement à ce qui se passera dans l'expérience 3 de 1971 et dans l'expérience « grandeur nature » de 1973. Les sujets avaient pour instruction d'évaluer 27 paris différents dont l'utilité attendue variait entre $-3,00\$$ et $+3,00\$$ ⁴. Il y avait un groupe qui devait évaluer l'attractivité des paris sur une échelle de satisfaction graduée de -5 à $+5$. Un autre groupe, sous divisé en trois classes avait pour instruction de donner une évaluation monétaire selon trois méthodes différentes (une méthode par classe) :

1 – *Premier sous-groupe* : déterminer le montant le plus important qu'ils seraient prêts à payer pour jouer. Pour les paris avec une utilité attendue négative, les sujets devaient déterminer quelle rétribution il faudrait leur octroyer pour jouer le pari.

2 – *Deuxième sous-groupe* : un ticket de loterie était distribué pour chacun des paris et les expérimentateurs demandaient quel est le prix le plus bas auquel les sujets accepteraient de vendre leur ticket. Pour les billets avec une utilité attendue négative on demandait aux sujets quel est le prix le plus bas qu'ils accepteraient de payer pour ne pas jouer le pari.

3 – *Troisième sous-groupe* : déterminer un prix équitable pour chaque pari, sans savoir si c'étaient eux ou les expérimentateurs qui allaient jouer les paris.

¹ L'article de 1968 décrit trois types d'expériences différentes. C'est seulement dans le cadre des expériences de 1971 que les paris seront décrits par les concepts de « Pbet » et « \$bet », c'est-à-dire une probabilité élevée de gagner un gain modeste pour le premier, une probabilité faible de gagner une somme plus importante pour le second. Voir L/S (1971), p 53.

² Slovic/Lichtenstein (1968), p 43. Il y a donc quatre possibilités pour les deux paris.

³ Idem, p 42.

⁴ Idem, p 44, pour le tableau.

Nous remarquons donc que nous avons deux types d'instructions différentes déclinées sous plusieurs modes : évaluer en terme d'envie de jouer (« rating »), proposer un prix, une enchère (« bidding »). Il s'agit d'étudier quel impact ces différentes instructions ont sur les sujets. Si nous mettons en rapport ces expériences avec la théorie N/M, nous voyons que nous avons affaire à des tâches plus complexes que celles décrites par la TUA, dans laquelle on demande au sujet d'ordonner les options et d'élire celle qui a l'utilité la plus élevée. Et ceci pourrait nous amener à penser que ces expériences surdéterminent empiriquement la TUA, puisqu'elles impliquent des tâches que cette dernière n'avait pas formalisée.

Une autre remarque s'impose c'est qu'il s'agit de protocoles expérimentaux très élaborés, les plus élaborés que nous ayons vus jusqu'à présent. Ils sont dans l'esprit de ce qu'Edwards envisageait dans son article de 1954 et de ce que lui-même a pratiqué par ailleurs. Par exemple les résultats sont analysés avec des méthodes statistiques élaborées comme le χ^2 .

Les auteurs avaient eu eux aussi, à l'instar d'Edwards, l'idée d'introduire un coefficient de pondération (ω_n) pour chaque dimension de risque. Ces coefficients servaient à mesurer l'importance des différentes dimensions et la façon dont les sujets se les représentaient. Ainsi l'attractivité d'un pari pouvait être formalisée ainsi :

$$A(G) = \mu + \omega_1 P_g + \omega_2 \$g + \omega_3 P_p + \omega_4 \$p^1.$$

Comme le remarquent les auteurs, le résultat le plus important de ces expériences est que : « For the groups as a whole, ratings correlated more highly with P_w than with any other dimension, whereas bids correlated with $\$l$ »². Ils avaient donc mis en avant le fait que selon la tâche demandée, ce n'est pas le même stimulus qui retient l'attention. Ce qui signifie d'une façon plus générale que la décision est tributaire de l'information³, comme l'avait déjà soutenu Simon (1955), Edwards (1961) et Ellsberg (1961). Nous avons donc des sujets qui classent des paris sur une échelle subjective en fonction de la probabilité de gagner, mais qui estiment

¹ Nous ne sommes pas très loin de la façon dont Kahneman et Tversky formaliseront une perspective normale dans la Théorie de la Perspective. Voir Kahneman/Tversky (1979), p 31.

² Slovic/Lichtenstein (1968), p 46.

³ Il y a une sélection de l'information par le sujet selon la tâche demandée, dans le cadre d'une évaluation monétaire c'est la perte possible qui est retenue comme essentielle, alors que dans l'évaluation en terme de volonté de jouer c'est la probabilité de gagner. Il y a donc une dissymétrie évidente dans les effets qu'ont les différentes dimensions sur le sujet. Cette dyssymétrie devra être analysée de plus près.

financièrement ces mêmes paris en fonction de la somme à perdre ou à gagner. C'est cela même qui produit le renversement des préférences. Par ailleurs ces résultats seront confirmés par les expériences de 1971 et 1973, de façon massive.

L'idée centrale de l'article de 1971 : « Reversals of Preferences between Bids and Choices in gambling Decisions »¹, est articulée autour de trois expériences différentes ayant pour but d'approfondir les résultats obtenus en 1968 concernant les dimensions qui interviennent dans la construction des préférences subjectives. Comme nous le notions plus haut, l'idée est d'élaborer des paris de deux sortes : les paris P avec une large probabilité de gagner une petite somme, par exemple .99 de gagner 4\$, et .01 de perdre 1\$; et les paris \$ avec une petite probabilité de gagner une somme plus importante, par exemple .33 chances de gagner 16\$ et .62 de perdre 2\$². Armés de ce type de paris on demande aux sujets – comme dans l'article de 1968 – tout d'abord de choisir quels paris ils préféreraient *jouer*, et ensuite on leur demande de donner une *estimation financière* des paris.

Ce que les trois expériences³ ont mis en avant de façon massive c'est que les sujets préféreraient jouer les paris P mais attribuaient une valeur monétaire supérieure aux paris \$: « In three experiments, Ss frequently chose one bet from a pair of bets and subsequently bid more for the bet they did not choose »⁴. Cela implique que selon la tâche qui est demandée au sujet (choisir/évaluer) d'autres processus sont à l'œuvre que la simple considération de l'utilité attendue. Ce phénomène de renversement des préférences a un impact très important⁵ sur la TUA et la théorie de la décision : « These reversals clearly constitute inconsistent behavior and violates every existing theory of decision making »⁶.

Ce que nous enseignent ces expériences c'est que tout d'abord les probabilités ne sont pas suffisantes comme source d'information dans la détermination de l'utilité. Le sujet va être attentif à d'autres facteurs, d'autres sources d'information surtout si on leur demande des tâches différentes (choisir/évaluer)⁷.

¹ Il est important de remarquer que si l'article de 1968 n'était pas destiné à tester les aspects normatifs et prédictifs de la TUA, l'article de 1971 commence lui par ce problème.

² L/S (1971), p 53.

³ Dans ces trois expériences, seuls les paris de la troisième sont joués réellement par les sujets.

⁴ Idem, p 63.

⁵ Les expériences de 1973, en faisant sortir le RP du laboratoire allaient accréditer encore plus la robustesse du phénomène : « There is a natural concern that the results of any experiment may not be replicated outside the confine of the laboratory ».

⁶ L/S (1971), p 63.

⁷ Slovic/Lichtenstein (1968), p 46.

Avec le phénomène du PR c'est l'axiome d'ordonnancement qui est violé de façon directe, mais, l'axiome d'indépendance est lui aussi touché de façon indirecte puisque lorsqu'on demande aux sujets d'évaluer monétairement, l'aspect probabiliste de la perspective n'est même pas pris en compte, et cela au profit de la somme à perdre ou à gagner. Comme le notent L/S :

« The finding that P_w has more import than P_l is particularly critical because it constitutes a violation of the assumption in the SEU model, that there is just a single probability function $S(P)$, which is independent of the payoffs »¹

Ces expériences, et la façon dont elles ont été construites nous montrent qu'une perspective est une entité complexe, que le sujet n'analyse pas d'un bloc, mais qu'il divise son attention et son intérêt entre les différentes *dimensions* qui constituent cette alternative. Ces analyses nous montrent aussi comment ces différentes dimensions interagissent les unes avec les autres, ou pas, comme c'est le cas par exemple entre les probabilités et les gains dans la tâche d'estimation financière. Nous pourrions dire que ces dimensions sont tout à fait irréductibles : elles ont chacune un effet spécifique sur la construction des préférences. Et encore ici nous situons-nous dans un contexte particulièrement épuré où seules quatre dimensions entrent en ligne de compte.

Que dire alors de situations dans lesquelles les dimensions se trouvent pour ainsi dire démultipliées. Comme le remarque L/S, l'image du décideur qui émerge ici est celle d'une personne : « struggling to integrate several sources of information into a single choice or judgment »². Nous verrons dans notre troisième chapitre que c'est là une image du décideur qu'il conviendra d'affiner, notamment en ceci qu'une théorie de la décision doit prendre en compte tous les types de décisions, et couvrir l'échelle graduée qui va de la décision presque inconsciente aux décisions exceptionnelles³. Nous verrons que le nombre de dimensions qui interviennent peut être sensiblement augmenté par rapport à ce que proposent L/S, mais qu'il peut être aussi diminué.

¹ Slovic/Lichtenstein (1968), p 49.

² Slovic/Lichtenstein (1968), p 50

³ Voir notre schéma, p 219.

Nous voilà arrivés au seuil du moment central de notre travail qui va être consacré à l'analyse de la Théorie de la Perspective, élaborée par Kahneman et Tversky en 1979. Depuis le paradoxe d'Allais, les charges contre la TUA se sont accumulées et aggravées, ce qui a permis l'émergence d'éléments conceptuels importants pour la suite de nos analyses du concept de décision. Quelques remarques synthétiques s'imposent ici. Tout d'abord nous avons pu nous rendre compte que c'est l'axiome d'indépendance qui sert le plus souvent d'angle d'attaque aux critiques, directement comme chez Allais (1953) et Ellsberg (1961), ou indirectement comme chez L/S. L'axiome d'indépendance est le centre de gravité de la TUA¹, il était donc important de montrer sa défaillance face à l'expérience, et de montrer que l'indépendance postulée des probabilités et des gains est une illusion. Il faut ménager un rôle à la perception du risque qui doit aller au-delà de la simple concavité de la fonction d'utilité, il faut montrer comment le risque est lui-même un facteur de pondération décisionnel, ce que la TUA ne faisait pas. D'autre part, de tous les axiomes que nous avons étudiés et qui constituent le cœur de la TUA, seule la propriété archimédienne semble avoir été épargnée par les critiques. Comme nous le remarquons – et ce sera à approfondir –, il appert que cette propriété a une portée normative non négligeable, et qu'elle pourra sans doute jouer un rôle dans notre propre théorie. Un rôle important puisque comme nous le remarquons dans notre analyse de cet axiome, nous pouvons vérifier son aspect descriptif, prédictif et normatif quotidiennement.

Les travaux d'Allais, mais encore plus ceux d'Edwards, font référence de façon explicite à la formalisation d'un coefficient de pondération sur les probabilités qui devrait représenter la distorsion subjective des probabilités objectives. La Théorie de la Perspective sera la formalisation explicite de ce type de déformation, de distorsion. Nous ferons un usage renouvelé du concept de coefficient de pondération dans notre troisième chapitre. En effet nous montrerons que des coefficients de pondération sont à l'œuvre dans d'autres facteurs que les probabilités², comme le temps par exemple.

L'article d'Ellsberg a appelé notre attention sur le concept de contexte décisionnel, qui devra être remplacé par celui d'environnement décisionnel, en y intégrant d'autres

¹ Voir Fishburn (1970), p 108 ; Fishburn/Wakker (1995).

² Probabilités dont nous pensons qu'elles sont très loin d'être utilisées systématiquement. Sur l'échelle du décidable, seulement une partie fait appel au calcul des probabilités.

informations que les probabilités. D'un point de vue plus général, la redéfinition du contexte décisionnel et son élargissement à celui d'environnement nous feront sortir des limites de la Théorie de la Perspective qui utilise le concept classique de risque dans l'article de 1979, et le concept classique d'ignorance dans la Théorie Cumulative de la Perspective de 1992. Comme nous le remarquons plus haut, il conviendra de faire la part entre les décisions qui font appel à des probabilités objectives ou distordues, et les décisions qui reposent sur des critères tout à fait autres.

Mais il ne faudra surtout pas oublier que l'environnement décisionnel qui remplacera le contexte décisionnel n'aura de sens que s'il est intégré au travail du SRD. Système dynamique et évolutif à partir duquel le sujet décidera en fonction de ses affects, de ses valeurs, de ses croyances, et de tout ce qui fait son irréductibilité. Ce système de référence, qui sera introduit dans le troisième chapitre sera pensé sur le modèle de la fonction de valeur de la Théorie de la Perspective. Nous aurons aussi l'occasion de montrer que c'est précisément à partir de ce système que s'opère les différentes pondérations dans les processus décisionnels, il s'agira de fonder la théorie de la décision sur ce socle subjectif dont la fonction principale est l'assimilation au sens strict du terme, c'est-à-dire « rendre semblable à soi », il s'agira donc de montrer qu'il n'y a pas de décision sans représentation subjective.

C – La Théorie de la Perspective.

1) les heuristiques et les biais.

Il pourrait sembler paradoxal de débiter un chapitre consacré à la Théorie de la Perspective¹ par une analyse de l'article de 1974 : « judgment under uncertainty : Heuristics and biases »² cependant deux raisons autorisent ce choix. Tout d'abord la théorie des heuristiques et des biais (HB) vient en quelque sorte continuer sur autre plan la série de critiques, issues du champ de l'expérience, entamées par Allais. Comme nous le verrons, le fait que les sujets utilisent de façon spontanée un certain nombre d'heuristiques va contre l'esprit même de la TUA, ce fait indique que les sujets ne calculent pas.

D'autre part, et c'est peut-être moins évident³, il semblerait que la théorie des HB prépare et anticipe ce qui dans la TP deviendra la phase d'édition⁴, dans laquelle le sujet recompose et reconstruit en quelque sorte les alternatives qui lui sont présentées. Nous faisons ici l'hypothèse qu'il y a une analogie et une solution de continuité entre les deux théories, et que certaines « méthodes » que les sujets utilisent pour reformuler les perspectives peuvent tout à fait être considérées comme des heuristiques telles que celles analysées dans l'article de 1974.

L'introduction du bayésianisme dans le domaine de la psychologie de la décision, par Edwards/Lindman/Savage (1963), allait donner un point d'ancrage normatif à l'analyse des décisions quotidiennes, et du jugement intuitif⁵ : un terme de comparaison. Cependant les

¹ TP à partir de maintenant. Toutes les citations de l'article fondateur de 1979 proviennent du volume préparé par Kahneman et Tversky (K/T), et publié en 2000 sous le titre : *Choices, Values, and Frames*. Amos Tversky est mort en 1996, il n'a donc pas vu l'ouvrage achevé. L'article sur la TP se situe aux pages 17- 43. Nous citons K/T (1979).

² *Science*, 1974, 185, 1124-1131. Nous citons la réédition qu'en ont fait Kahneman, Slovic et Tversky dans le volume : *judgment under uncertainty : Heuristics and biases*. Paru en 1982, l'article se situe aux pages 3-20. Pour une très bonne introduction historique, voir Gilovitch/Griffin (2002), p 1-18 ; Keren/Teigen in Koehler/Harvey (Eds.) (2004), p 89-110 ; Gigerenzer in Koehler/Harvey (Eds.) (2004), p 62-89 ; Manktelow (2012), p 3-39.

³ D'autant moins évident que, hormis Keren/Teigen (2004), ce fait ne semble avoir été soulevé par personne à notre connaissance.

⁴ K/T (1979), p 28-30.

⁵ « Intuitive Judgment », le terme est peut-être mal choisi du fait de la surdétermination du concept d'intuition dans la philosophie, surdétermination qui n'a d'ailleurs permis à ce jour aucune définition réellement consensuelle. Il aurait peut-être mieux valu parler de « jugement quotidien », « jugement spontané », ou « jugement habituel », nous ne sommes pas à proprement parler ici dans le domaine de la philosophie ou de l'épistémologie, même si certaines composantes épistémiques ou doxastiques entrent en jeu de façon assez naturelle.

résultats des expériences allaient très vite montrer deux types de « défaillances »¹ de la part des sujets :

a – Un décalage très net entre les prévisions théoriques et ce que les sujets font dans la réalité. Encore une fois l’aspect normatif des théories en présence allait être mis en défaut, et au premier chef, le bayésianisme, soutenu, entre autres par Edwards et Savage

b – Comme le remarquent K/T, les expériences furent : « the demonstration of a substantial discrepancy between the objective record of people’s success in prediction tasks and the sincere beliefs of these people about the quality of their performance »².

D’autre part, les travaux de Simon (1955) et (1956) avaient mis l’accent sur le fait que nous faisons un usage seulement « écologique » et limité de la rationalité³, c’est-à-dire que les critères auxquels auraient dû se soumettre les sujets pour se conformer à la théorie de la décision étaient irréalistes. La théorie de la décision, aussi achevée soit-elle, n’est que très peu descriptive, et encore moins normative. L’hypothèse des HB que construisent Kahneman et Tversky aura le mérite de proposer un modèle descriptif de la façon dont les sujets jugent et décident réellement.

Une troisième raison qui motive l’étude de ces phénomènes dans ce travail est le rôle que le concept d’heuristique va être amené à jouer dans notre troisième chapitre. Aussi posons-nous la question immédiatement : qu’est-ce qu’une « méthode heuristique » en théorie de la décision ? Comme le soulignent K/T : « People rely on a limited number of heuristics principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations »⁴. Cette citation nous apprend d’emblée deux choses

¹ Si nous parlons ici de « défaillances », c’est parce que nous nous situons au point de vue de K/T qui voyaient dans ces résultats une forme de « misère cognitive ». Nous n’assumons pas ce point de vue et nous rangeons plutôt à celui soutenu par exemple dans Cosmides/Tooby (1996) ; et Gigerenzer (1991), (2008), et (2011).

² Kahneman/Slovic/Tversky (1982), p. XII. Il y a donc un double décalage : par rapport à une norme et par rapport à l’autoévaluation subjective.

³ Rationalité écologique et rationalité limitée seront deux concepts sur lesquels nous reviendrons au début de notre troisième chapitre, ils seront tous deux consacrés par les travaux de Gerd Gigerenzer dans ce domaine. Voir à ce propos ses ouvrages de 2002 et 2008, ainsi que l’imposant volume paru en 2011 sous le titre : *Heuristics. The foundations of adaptive behavior*.

⁴ Tversky/Kahneman (1974), p. 4.

importantes sur les heuristiques : elles ont pour vocation de simplifier¹ une tâche cognitive ou une situation perçue comme complexe, par ailleurs elles s'appliquent dans le cadre des évaluations probabilistes et des jugements qui sous-tendent les processus décisionnels. Précisons. Étymologiquement le terme d'« heuristique » désigne en grec « trouver un moyen de »² ; à l'époque contemporaine, et notamment depuis son utilisation par Polya en 1945 dans le domaine des mathématiques³ le concept désigne un moyen pour résoudre un problème particulier en utilisant une règle – construite empiriquement – qui n'est pas un algorithme ni une méthode de décision déterminée⁴. Polya oppose heuristique à analytique, la première méthode étant à verser au registre des méthodes provisoires en attente de mieux. Plus vulgairement on pourrait associer heuristique et « recette », « tour de main ».

Dans le domaine que nous étudions, les heuristiques seront des formes de *raccourcis décisionnels* qui permettront certes de répondre plus vite⁵ à la situation, mais qui risquent néanmoins de conduire à des erreurs systématiques, des déviations, autrement dit des biais. En effet comme nous allons le voir, les heuristiques peuvent produire une transformation des données originelles qui n'est pas sans conséquences⁶. La question qui émerge est de savoir si ces heuristiques sont, pour le sujet qui décide, un atout ou un handicap ? Les choses ici doivent être nuancées. À l'origine Kahneman et Tversky voyaient les HB comme une défaillance des sujets par rapport à des théories formelles comme les statistiques ou les probabilités par exemple. Théories, qui dans l'esprit des auteurs avaient certaines conséquences normatives. C'est d'ailleurs la définition exacte d'un biais : la déviation par rapport à une norme établie. Et dans notre contexte, *les biais sont les effets des heuristiques*.

Cependant tout un courant de pensée initié par Simon (1955), Cosmides/Tooby (1996),

¹ On voit donc ici le lien immédiat qu'entretient ce concept avec la phase d'édition de la TP.

² La racine est le verbe grec *heuriskô*, d'où dérive aussi le célèbre « Euréka » Archimédien. Voir dictionnaire *Lacroix-Magnien*, p 738.

³ Einstein utilise le terme dans un sens assez similaire dans un de ses articles de 1905 : « un point de vue *heuristique* concernant la production et la transformation de la lumière ». C'est nous qui soulignons.

⁴ Il s'agit en fait chez Polya de construire la règle en fonction du problème. Pour de bonnes définitions des concepts d'heuristiques et de biais, voir Baron (1988), p 146-159 ; Keren/Teigen in Koehler/Harvey (Eds.) (2004), p 91-93 ; plus spécifiquement pour le concept d'heuristique et son histoire, voir Gigerenzer/Todd/ABC Research Group (1999), p 25-29 ; et Gigerenzer/Brighton (2009), p 2-3. K/T ne proposent pas à strictement parler une définition générale approfondie de ces concepts, les définitions émergent au cours des analyses et sont proposées in situ. Les deux volumes susmentionnés explorent le phénomène des HB dans toutes les dimensions possibles et apportent de nombreux éléments définitionnels en fonction des contextes dans lesquels les phénomènes sont étudiés.

⁵ Voire même spontanément, ce qui pose une autre forme de problème, débattu en psychologie cognitive aujourd'hui : l'application des heuristiques est-elle automatique ? Voir Frederick (2002).

⁶ Il sera donc important pour nous d'analyser le lien entre les heuristiques et le concept de pondération.

et Gigerenzer (1991), soutiendra au contraire que les heuristiques sont un atout, puisqu'elles permettent une réponse immédiate, pertinente et écologique à toutes sortes de situations décisionnelles¹.

L'interrogation de KT dans cet article va porter sur la façon dont les heuristiques « canoniques » vont produire de façon systématique une série de biais cognitifs, ce qui signifie que pour une heuristique déterminée nous aurons différentes sortes de biais qui seront causés par elle, et qui auront en commun cette caractéristique de violer un principe, une loi ou un axiome d'une théorie établie. D'une certaine façon nous pouvons dire, et c'est ce que font K/T que *le biais est à la psychologie ce que l'illusion est à la perception*. Il est donc clair que dans l'esprit des auteurs, le but était le suivant : déterminer de façon précise quels sont les différents types d'heuristiques et de biais afin de pouvoir les éviter et les prévenir. Pour K/T les biais sont les marques d'une défaillance cognitive. Contrairement à ce que soutiendra Gigerenzer quelques années plus tard, K/T pensent qu'il est mieux pour les sujets d'éviter les heuristiques puisqu'elles sont sources d'erreurs de jugement et d'évaluation.

L'article se déploie autour de l'analyse de trois heuristiques principales : La représentativité (« representativeness »), la disponibilité (« availability »), l'ajustement et l'ancrage (« adjustment and anchoring »)². Nous allons ici analyser la structure de chaque heuristique³ et présenter les biais qui nous semblent avoir le plus d'importance pour la suite de notre travail, c'est-à-dire ceux qui sont le plus susceptibles de jouer un rôle dans la TP et plus généralement en contexte décisionnel.

Il faut remarquer que chaque biais est lié à un élément de la théorie des probabilités ou des statistiques, qu'il néglige ou duquel il dévie. Ce qui nous fait revenir sur la question que nous nous posions plus haut, dans ce travail, à savoir dans quelle mesure les sujets sont-ils capables de tirer d'une situation donnée l'information statistique pertinente, et

¹ Voir à ce sujet, Gigerenzer (1996), et Gigerenzer/Brighton (2009). On parle de rationalité écologique lorsque la réponse à une situation est fonction de l'environnement décisionnel. Du point de vue de Gigerenzer, les heuristiques sont écologiques parce qu'elles exploitent la structure de l'environnement.

² Voir Tversky/Kahneman (1974), p 4-18.

³ Comme le montre le deuxième volume consacré aux HB, édité par Gilovich/Griffin/Kahneman en 2002, le nombre d'heuristiques n'a pas sensiblement augmenté, et les recherches sont restées dans le cadre des phénomènes présentés en 1974. Ce sont les directions théoriques qui ont évolué, notamment vers une réflexion sur l'automatisme à l'œuvre dans les heuristiques conduisant les chercheurs à s'interroger sur la possibilité de deux systèmes de cognition, l'un quasi automatique, l'autre réflexif. Les heuristiques seraient alors à verser au registre du premier système, voir Kahneman/Frederick (2002) ; Sloman (2002) et le livre de Kahneman (2010). Il en va autrement des biais dont le nombre est allé croissant, Baron (1988), p 56-57, en dénombre plus de vingt divisés en cinq types.

dans quelle mesure sont-ils capables de l'utiliser ? Gigerenzer/Goldstein (1996) radicaliseront le point de vue avec sa théorie du comportement heuristique : les sujets n'ont tout simplement pas besoin de la totalité de l'information.

Le premier type d'heuristique analysé par K/T est la représentativité qui a pour vocation de faciliter la réponse aux questions suivantes : quelle est la probabilité que A appartiennent à la classe B ? Quelle est la probabilité que du processus B puisse émaner l'événement A ? Quelle est la probabilité que B engendre A ? Pour les auteurs :

« ... probabilities are evaluated by the degree to which A is representative of B , that is, the degree to which A resembles B . For example, when A is highly representative of B , the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B , the probability that A is judged to be low »¹.

Le principe général de cette heuristique² repose sur le concept de *ressemblance*, par exemple entre un objet et la catégorie à laquelle il pourrait appartenir. Les sujets vont juger en considérant que la ressemblance et la représentativité sont les seules informations pertinentes et négliger le reste de l'information disponible, par exemple les probabilités premières et la taille de l'échantillon. L'heuristique de représentativité est d'une part liée au concept de stéréotype³, d'autre part à la sélection de l'information pertinente, c'est-à-dire celle qui demandera le moins d'effort. Elle donne lieu à pas moins de six biais cognitifs différents : insensibilité aux distributions a priori de probabilités, insensibilité à la taille de l'échantillon, mauvaise compréhension du concept de chance, prédictions biaisées par la description actuelle⁴, illusion de validité, mauvaise conception de la régression vers la moyenne. L'insensibilité aux distributions a priori peut être analysée grâce à l'exemple suivant⁵ : les

¹ Tversky/Kahneman (1974), p 4.

² Cette heuristique et les biais qui lui sont associés seront retravaillés dans Kahneman/Frederick (2002), notamment à la lueur de l'hypothèse des deux systèmes cognitifs.

³ Voir Baron (1988), p 149. Les concepts de ressemblance, représentativité et stéréotypes sont eux-mêmes très liés à ce qu'on appelle « opinion » dans la philosophie, depuis Platon.

⁴ Il s'agit ici en fait de fonder une prédiction sur la description actuelle sans prendre en compte d'une part la fiabilité de l'information, et d'autre part le fait que l'information peut évoluer avec le temps. Par exemple les sujets auront tendance à faire une prédiction favorable sur le long terme pour une entreprise qui fait actuellement un profit élevé, sans tenir compte du fait que l'information est peut-être erronée, et que les choses peuvent changer avec le temps. Il s'agit d'utiliser la connaissance présente, même déficiente, et d'en faire un usage prédictif.

⁵ Tiré de K/T (1973), p 53. Il s'agit d'un article central quant au contenu empirique et sur la représentativité, il contient des considérations méthodologiques importantes, notamment p 63-65.

auteurs donnaient aux sujets la description de quelques individus pris au hasard dans un groupe de 100 professionnels (ingénieurs et avocats). Ils demandaient alors aux sujets de déterminer, pour chaque description, la probabilité qu'elle soit celle d'un avocat ou d'un ingénieur. Dans une expérience on donnait aux sujets les distributions de probabilités suivantes : 70 ingénieurs, 30 avocats, et dans une deuxième, à d'autres sujets, la distribution inverse. Or, dans les deux cas, les sujets donnèrent les mêmes jugements probabilistes, négligeant ainsi les distributions a priori de probabilités et le théorème de Bayes¹. Leurs jugements étaient établis en fonction de la ressemblance de la description avec le stéréotype d'une classe ou de l'autre².

Dans la même veine que le biais précédent, la représentativité est aussi la cause du fait que les sujets ignorent la taille de l'échantillon duquel les éléments à évaluer sont tirés. Par exemple si on demande aux sujets lequel de deux hôpitaux est plus susceptible de voir son taux de naissances de garçons ou de filles varier de façon significative, en sachant qu'il y a un petit et un grand hôpital, on s'aperçoit que les sujets attribuent un taux de variation identique aux deux établissements. Or, les statistiques nous apprennent que les variations seront plus sensibles dans le petit hôpital³. Il appert ici que les sujets sont sensibles aux proportions qui décrivent l'échantillon, à savoir ici que le taux des naissances de garçons ou de filles est à peu près égal à 50 % en règle générale, mais qu'ils sont insensibles à la taille des échantillons (petit hôpital/grand hôpital)⁴.

Un troisième biais lié à la représentativité consiste à voir dans la chance et dans les suites aléatoires un processus auto correcteur⁵. C'est le fameux « sophisme du joueur »⁶ qui

¹ Comme le remarquent K/T, les probabilités a priori sont considérées seulement si elles sont la seule information disponible dans le protocole, dès qu'une description – même moyennement informative – est rajoutée elle remplace immédiatement les considérations probabilistes. C'est donc le rapport au stéréotype qui a le plus d'impact.

² Voir Manktelow (2012), p 16.

³ Tversky/Kahneman (1974), p 6-7. L'exemple est bien analysé dans Hacking (2004), p 82 ; Manktelow (2012), p 17. Les statistiques nous indiquent que plus l'échantillon est petit et plus les variations sont sensibles. Un autre exemple bien connu de ce type de biais se trouve dans Tversky/Kahneman (1982), il s'agit du « taxicab problem », il est lui aussi analysé dans les deux ouvrages cités en début de note.

⁴ En fait si nous regardons les choses d'un peu plus près, nous voyons que les sujets sont victimes de ce que K/T appellent la « croyance en la loi des petits nombres ». C'est d'ailleurs le titre de leur article de 1971 qui visait à montrer que les sujets – ici des scientifiques – attribuent la plupart du temps le même effet à un petit et à un grand échantillon. Un scientifique victime de la croyance en la loi des petits nombres « will have exaggerated confidence in the validity of conclusions based on small samples », Tversky/Kahneman (1971), p 25. Il s'agit d'une forme de biais de confirmation.

⁵ Un processus dans lequel « a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium », Tversky/Kahneman (1974), p 7.

⁶ Voir Baron (1988), p 152 ; Andler (2004), p 373, l'excellente analyse de Hacking (2004), p 33-45 ; Manktelow (2012), p 9 et 199 ; Gigerenzer/Brighton (2009), p 18, font remarquer que le même processus est à l'origine du sophisme du joueur et du sophisme de la bonne main. Le sophisme de la bonne main, issu du domaine sportif, mais pouvant être généralisé à d'autres

peut avoir des conséquences assez importantes sur nos décisions. Il s'agit par exemple, dans un jeu de roulette, de croire qu'après une suite de 7 noirs, le rouge va nécessairement sortir, car la suite de 7 noirs n'est pas *représentative* de ce que doit être une suite aléatoire (communément vue comme une alternance approximativement régulière de noir et de rouge). Ce sophisme repose sur deux représentations biaisées : les probabilités et la loi des grands nombres. Les sujets ont tendance à oublier que par exemple si je tire quatre 6 d'affilés avec un dé non pipé, la probabilité de faire un autre 6 reste 1/6, les sujets ont en fait l'impression que la probabilité diminue, car elle dépend des lancers précédents, or ils sont tous indépendants du point de vue probabiliste¹. C'est-à-dire que le tirage présent est influencé par les tirages passés. Par ailleurs les sujets ignorent aussi le fait que ce n'est que sur une série très longue de lancers que la probabilité s'approchera de plus en plus de sa valeur moyenne de 1/6.

Le sophisme du joueur a ceci de particulier qu'il peut, en contexte décisionnel, conduire les sujets à un certain optimisme et à se doter d'une capacité de prédiction fictive. En effet, si dans une situation où intervient un élément comme le tirage au sort répété, les sujets commettent cette erreur ils peuvent se croire justifier non seulement à « parier » de façon irrationnelle, mais aussi de prédire l'événement qui aura lieu. Ce qui peut avoir de fâcheuses conséquences.

L'heuristique de disponibilité repose elle sur un mécanisme différent : « There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind »². Il s'agit ici de relier les décisions des sujets à l'information stockée dans leur mémoire. Ce lien entre mémoire et décision est important, et nous y reviendrons quand nous aborderons le problème de l'habitude dans notre troisième chapitre, en particulier lorsque nous étudierons la contextualisation³. D'autre part, ce que nous indique ce phénomène, et ce que vont confirmer les biais qui en découlent, c'est que la plupart du temps les sujets ne font pas d'effort réel pour acquérir une

contextes, consiste à dire que si un basketteur vient de réussir une série de x lancers francs, il va nécessairement réussir le $x+1$ ème. On voit ici qu'il s'agit exactement de l'opposé du sophisme du joueur qui dit que la suite ne doit pas continuer mais changer. L'exemple sert aux auteurs à discréditer la représentativité comme explication, en effet dans ce cadre là, le même concept servirait à décrire deux phénomènes opposés.

¹ Voir Hacking (2004), p 33-47.

² Tversky/Kahneman (1974), p 11. Structurellement cette heuristique est assez proche de l'heuristique de reconnaissance que les sujets utilisent régulièrement et qui est exposée par Gigerenzer/Goldstein (2009) in Gigerenzer (Ed.) (2011), p 58-82.

³ Voir le Schéma p 219. L'opération de contextualisation permet de mesurer le rôle que peut jouer la mémoire dans les différentes situations décisionnelles qui peuvent se présenter. Plus les situations seront habituelles et non critiques et plus la mémoire jouera un grand rôle.

information nouvelle ou plus précise sur l'alternative proposée, ils s'en tiennent donc aux éléments qu'ils peuvent mobiliser immédiatement, utiliser le plus rapidement et qui seront selon eux les plus efficaces¹. On pourrait donc voir ici – du point de vue de K/T – une forme de paresse intellectuelle, paresse qui pourrait par exemple provenir du fait que les sujets ont un *système de référence décisionnel stable, sédimenté autour d'habitudes élémentaires*, et que bien qu'il y ait possibilité pour ce système d'évoluer durant le cours de l'existence, les sujets ont tendance à assimiler ce système de référence au statu quo². Cela induit donc une forme de paresse dans la quête de l'information, pour des décisions assez élémentaires³.

La disponibilité est elle aussi la cause de plusieurs biais, nous détaillerons seulement les deux suivants : biais dû à la facilité de retrouver des éléments connus dans un ensemble, biais dû à l'efficacité de l'ensemble dans lequel on cherche l'information. Deux autres biais sont produits par cette heuristique, mais ils sont moins liés à la décision, nous les laisserons donc de côté : biais lié à « l'imaginabilité », et biais liés aux corrélations illusives. Remarquons seulement que ces deux derniers phénomènes ne sont plus reliés à la mémoire des sujets.

Le premier biais se présente comme suit : l'évaluation de la taille de deux ensembles identiques sera faussée si les éléments de l'un sont bien connus par le sujet, et les éléments de l'autre moins. Quand nous parlons ici de connaissance, nous entendons par là seulement ce que les sujets ont à leur disposition dans leur mémoire de façon spontanée. K/T avaient fait l'expérience suivante⁴ : on faisait entendre aux sujets une liste de noms de femmes et d'hommes connus, dans l'une les hommes étaient légèrement plus connus que les femmes, et dans l'autre c'était l'inverse. Il y avait bien entendu le même nombre de membres des deux sexes dans chaque liste. Or les sujets évaluaient de façon systématique comme comportant plus d'éléments la classe dans laquelle se trouvaient le plus de noms connus. Ce qui signifie que les sujets s'en tenaient à l'information qu'ils pouvaient retrouver le plus vite dans leur

¹ C'est justement ici que vont se situer les divergences entre Gigerenzer et K/T. Pour le premier il s'agit de tirer le meilleur parti de l'information disponible, et il soutient même, à raison selon nous, que trop d'information, peut nuire à la prise de décision. Voir Gigerenzer/Brighton (2009), p 6-9.

² C'est justement cette caractéristique qui nous permettra de penser le système de référence décisionnel en nous appuyant sur la fonction de valeur en TP.

³ Ce point sera élaboré prochainement, à l'aide du concept de fonction de valeur. La question sera alors de savoir comment passer d'une théorie à caractère assez formel, à une théorie plus élaborée conceptuellement, c'est le pari qui est à l'origine même de ce travail. Quand nous parlons de décisions élémentaires, nous entendons par là des décisions qui reposent justement sur une pratique ou une habitude particulière. Nous parlons donc évidemment de décisions qui peuvent se répéter.

⁴ Elle est exposée en détail dans l'article de K/T (1972), p 175, qui est consacré à la disponibilité.

mémoire, et que seule cette information était jugée pertinente. Ce qui est reconnu aura donc toujours plus de poids que ce qui demande un effort pour aller chercher de l'information, il y aura donc une inertie très forte dans le système de référence décisionnel du sujet, ce qui pourra expliquer des concepts aussi importants que l'habitude et le statu quo ou immobilisme, qui mis en relation avec le concept d'aversion au risque pourrait expliquer une certaine forme d'akrasie.

Un autre biais, encore fortement lié à l'information stockée dans la mémoire des sujets consiste à répondre à une situation donnée, ou à un problème en se servant des informations qui viennent à l'esprit en premier lieu, sans chercher à enquêter plus loin. L'expérience était la suivante¹ : on demandait aux sujets de déterminer si dans un échantillon de mots anglais pris au hasard il était plus probable d'obtenir un nombre plus important de mots avec un *r* en première position, ou avec un *r* en troisième position. Or, comme ce sont les mots qui commencent par *r* qui viennent à l'esprit en premier, les sujets jugeaient que c'était eux les plus nombreux. Cependant, les entrées du dictionnaire indiquaient l'inverse.

Encore une fois ce biais nous indique une absence d'effort et un certain *automatisme*. Automatisme qui pourrait être expliqué par le fait que nous sommes souvent soumis à des listes classées alphabétiquement qui nous obligent à ne prendre en considération que les premières lettres des mots. Du point de vue décisionnel, ce phénomène vient confirmer ce que nous disions plus haut : *il y a une inertie forte du système de référence décisionnel*, une résistance au changement qui est renforcée par certaines habitudes intellectuelles, comme celles que nous venons d'étudier. La question est de savoir si cette inertie est utile ou nuisible ? En première approximation nous dirons qu'elle permet une cohérence en ce qui concerne les décisions quotidiennes, c'est ce qui fait que pour une certaine strate décisionnelle nous ne sommes pas toujours en train de changer d'avis.

La dernière heuristique est dite d'ancrage et d'ajustement ; elle est une anticipation du concept d'annulation dans la phase d'édition de la TP, mais elle anticipe aussi certains éléments que nous allons retrouver dans notre troisième chapitre, comme la possibilité pour la visée d'être biaisée. Cette heuristique produit trois biais dont deux sont importants pour nous. Son principe est le suivant :

¹ Pour plus de détails, voir K/T (1972), p 166-167.

« In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation »¹.

Le problème c'est que le point de départ de l'évaluation peut être estimé de différentes façons et conduire ainsi à différentes réponses face à une même situation selon la façon dont elle se présente². En règle générale l'ajustement du point de départ est insuffisant par rapport à la situation ou au problème, ce qui va mener à une sous-estimation. Un effet un peu plus pernicious est le suivant : la première lecture du problème par le sujet, spontanée et immédiate, va biaiser le point de départ de la résolution, mais plus encore, comme ce point de départ est lui-même biaisé, il va conduire à de fausses estimations basées sur lui, il s'agit donc en quelque sorte d'un biais de deuxième ordre. Comme le remarquent les auteurs, tout le processus peut être basé sur une « computation incomplète ». K/T ne définissent pas ce qu'ils entendent ici par *computation*, probablement ont-ils à l'esprit une forme de calcul élémentaire que les sujets utilisent de façon spontanée face à une information numérique. Inutile de dire que cette forme de calcul ne peut être que très approximative.

Le premier biais relié à cette heuristique présenté par K/T est celui de l'ajustement insuffisant, l'idée en est assez simple, et l'exemple qu'utilisent les auteurs est très clair : deux groupes de sujets devaient estimer le produit d'une même suite de chiffres avec un temps limité à cinq secondes, les premiers devaient évaluer le produit de la suite disposée dans cet ordre : $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, et les autres sujets devaient estimer le produit de la même suite en ordre ascendant. Dans les deux cas, comme l'ajustement est toujours insuffisant, les sujets ont proposé des sous-estimations : pour la suite ascendante les sujets proposaient un résultat moyen de 512, et pour la suite descendante un résultat moyen de 2250. Il s'agissait donc de deux sous-estimations puisque le bon résultat est 40 230. Cette différence dans l'évaluation peut être expliquée par le fait que dans le cas descendant comme les sujets ont un point d'ancrage plus haut, le résultat sera plus élevé que pour la suite ascendante qui commence par des chiffres plus modestes.

¹ Tversky/Kahneman (1974), p 14.

² Il s'agit ici de ce qui va devenir l'effet de cadrage ou l'effet du cadre de référence, dans lequel un même problème construit de différentes façons va conduire à des décisions différentes. Cet effet viendra accentuer le fait que les sujets ne calculent pas la valeur des perspectives mais se basent seulement sur l'information qu'ils peuvent tirer de façon immédiate et sans effort. Voir à ce sujet Baron (1988), p 267-271 ; K/T (1979), p 25-28 ; K/T (1986) ; Manktelow (2012), p 205-209.

Transposé dans le domaine de la décision, ce que fera la TP et l'article de 1986 sur l'effet de cadrage, ce biais peut avoir de lourdes conséquences, et notamment celle-ci : produire des choix différents pour des situations identiques construites et présentées dans des formes différentes. Cet effet viole le principe d'invariance en économie, c'est ce qu'avaient déjà mis en avant les expériences de L/S (1971), et ce que viendra confirmer la TP de 1979.

Un dernier biais important pour nous, lui aussi causé par l'heuristique d'ancrage et d'ajustement, est lié à l'évaluation des conjonctions ou des disjonctions d'événements. L'idée est la suivante¹ : on propose aux sujets trois types d'alternatives, un événement simple, un événement conjoint et un événement disjoint. Par exemple pour l'événement simple il s'agit de tirer une bille d'une urne contenant 50 billes rouges et 50 billes blanches. L'événement conjoint est le suivant : tirer une bille rouge 7 fois de suite – avec remplacement² – d'une urne contenant 90 billes rouges et 10 billes blanches. Enfin l'événement disjoint est présenté ainsi : tirer une bille rouge au moins une fois lors de sept tirages successifs – avec remplacement – d'une urne contenant 90 billes rouges et 10 billes noires. On demande ensuite sur quel événement ils préféreraient parier.

Du fait qu'ils prennent la probabilité de l'événement simple comme point d'ancrage dans l'évaluation des deux autres événements, et que cet ancrage est insuffisant, les sujets préfèrent l'événement conjoint à l'événement simple et ce dernier à l'événement disjoint, or les probabilités respectives de ces trois événements sont : .48, .50, .52. Les sujets choisissent donc l'alternative la moins favorable des trois. Les sujets surestiment la probabilité des événements conjoints et sous-estiment la probabilité des événements disjoints. Or le calcul des probabilités nous indique que la probabilité totale d'un événement conjonctif est inférieure à celle de l'événement simple³, et inversement pour les événements disjonctifs⁴. Comme le remarquent K/T ce phénomène, peut conduire les sujets à un optimisme et une confiance injustifiée quant à la réalisation d'un événement conjoint, et nous retombons un peu dans le travers que nous avons vu plus haut : si l'on croit A et B plus probable que A on prend beaucoup de risques en pariant sur la conjonction plutôt que sur l'événement.

¹ L'expérience n'a pas été menée par K/T, mais par Bar-Hillel (1973).

² Sur la différence très importante entre tirage avec remplacement et sans remplacement, voir Hacking (2004), p 47-66.

³ Il s'agit du fameux « sophisme de la conjonction », qui a fait couler beaucoup d'encre, voir entre autres Tversky/Kahneman (1982) et (1983) ; mais aussi la critique de Gigerenzer (1996).

⁴ Pour une approche très pédagogique de tous ces concepts et des notions fondamentales du calcul des probabilités, voir Baird (1992), p 182-200 ; Hacking (2004), p 67-123.

Pour conclure il convient de se poser un certain nombre de questions sur ces phénomènes. Tout d'abord demandons-nous dans quelle mesure les HB se placent en continuité avec les critiques que nous avons étudiées précédemment. Car nous remarquons tout de suite qu'il ne s'agit pas ici de réfuter tel ou tel axiome particulier de la TUA. De quoi s'agit-il alors ? Tout d'abord, comme c'était le cas chez L/S le but originel n'était pas de critiquer la TUA, mais de comparer les apports du bayesianisme avec le jugement quotidien, et aussi de tester la portée normative du bayesianisme. Bayesianisme qui repose sur l'hypothèse explicite que les sujets utilisent les probabilités subjectives dans l'élaboration et la révision de leurs croyances¹. Or il faut garder à l'esprit que depuis Savage, le bayesianisme s'était inscrit comme une complexification, un raffinement, et une *promotion intensifiée* de la TUA de N/M. Critiquer l'aspect normatif du bayesianisme c'était donc critiquer par ricochet l'aspect normatif de la TUA.

Par ailleurs, bien que n'étant pas une critique portant sur un axiome particulier, nous pouvons dire que les H/B critiquent en fait l'« esprit » de la TUA et du bayesianisme qui tous deux reposent sur cette hypothèse implicite que les sujets sont des calculateurs, avec les capacités de calcul adéquates². Les H/B nous montrent que ni les probabilités ni les statistiques ne sont pas maîtrisées par les sujets, comme le disent Kahneman et Tversky : « Statistical principles are not learned from everyday experience because the relevant instances are not coded appropriately »³. Autrement dit il n'y a pas un apprentissage spontané qui permettrait aux sujets de tirer l'information pertinente des situations données. Nous pourrions penser que cela est réservé aux spécialistes, mais avec cette limitation toutefois qu'eux aussi commettent des erreurs comme le montre l'article de 1971 sur la loi des petits nombres. Les HB préparent ainsi la critique de Gigerenzer, qui dans la continuité de Simon montrera que les réquisits concernant la rationalité sont placés trop haut. Le repositionnement de la rationalité par Gigerenzer dans le contexte de l'adaptativité constituera une avancée majeure vers un réalisme théorique concernant la décision humaine⁴. Réalisme théorique qui devra lui-même s'approfondir dans une *théorie de la volonté*, ce que nous ferons dans notre

¹ Voir à ce sujet, Oaksford/Chater (2007), et Chater/Oaksford (2008), p 3-25; ainsi que Griffiths/Yuille in Chater/Oaksford (Eds.) (2008), p 33-57.

² Cette idée sera critiquée assez rudement par Gigerenzer, voir entre autres Gigerenzer et Todd (1999), p 9.

³ Tversky/Kahneman (1974), p 19.

⁴ Voir les intéressantes remarques et l'intéressant schéma de Gigerenzer/Todd (1999) in Gigerenzer/Todd/ABC Group (Eds.) (1999), p 5-15. Mais d'une certaine façon, la TP elle aussi sera une avancée dans le réalisme, comme nous allons le voir.

troisième chapitre.

Deuxième question : quels liens entretiennent les HB avec la TP ? Comme nous le soulignons plus haut c'est un rapport d'analogie, car la théorie des HB prépare ce qui dans la TP deviendra la phase d'évaluation. Des opérations comme le codage ou l'annulation pourront être considérées comme des heuristiques telles que celles présentées dans l'article que nous venons d'étudier. Par ailleurs il conviendra de s'interroger justement sur le fait que cette phase d'évaluation ne recevra pas le même traitement systématique exhaustif que la phase d'édition. La réponse à cela est peut-être à chercher du côté de la théorie des HB.

2) La Théorie de la Perspective de Kahneman et Tversky (1979).

a – **Une critique renouvelée de la TUA.**

Comme nous l'avons vu dans les parties précédentes, dans le domaine de l'économie et de la psychologie un certain nombre de problèmes liés à la TUA avaient été très tôt mis en avant. Il s'agit notamment de la pondération des probabilités par les sujets¹, de la définition de l'utilité basée sur les gains et les pertes par rapport à l'avoir présent, et non plus sur l'avoir après maximisation². Enfin, la TUA avait été atteinte en son cœur même par les critiques successives de l'axiome d'indépendance qui depuis Allais n'avaient cessées de s'affiner et de s'aggraver. Par ailleurs, comme nous venons de le voir, la théorie des HB avait elle aussi enfoncé le clou, certes de façon plus indirecte dans le cercueil de la TUA. L'aspect normatif de cette théorie avait donc été sévèrement ébranlé dans différentes dimensions, ainsi que son aspect descriptif : le problème de l'achat simultané d'assurances et de tickets de loteries n'avait toujours pas pu être formalisé dans une seule théorie, et ce problème à lui seul hypothéquait sérieusement la pertinence descriptive de la TUA. Il fallait donc construire une théorie psychologique qui puisse intégrer ces contestations³ et les formaliser, une théorie

¹ Ce phénomène avait été originellement démontré par Preston/Baratta dans leur article de 1948, sur lequel s'appuie d'ailleurs Edwards (1954), et bien entendu par Allais (1951), suivi d'Ellsberg (1961) et L/S (1971). On peut penser cependant que le phénomène en question était « dans l'air » depuis Bernoulli et l'introduction de l'utilité attendue, ou de l'espérance morale chez Laplace. Voir notre introduction du chapitre I.

² Ici c'est Markovitz (1952), qui avait émis cette hypothèse, encore une fois elle sera relayée par Edwards (1954).

³ C'est-à-dire expliquer par exemple l'effet de rapport commun et de conséquence commune. Voir à ce sujet, Wu/Zhang/Gonzales in Koehler/Harvey (Eds.) (2004), p 404-407 ; et Sugden in Barberà/Hammond/Seidl (Eds.) (2007), p 694-699.

alternative, ce sera la Prospect Theory de 1979, élaborée par Daniel Kahneman et Amos Tversky¹.

Du point de vue de la philosophie, et il est important de le remarquer puisque Kahneman lui-même en fait état dans la préface de *Choices, Values, and Frames*, les problèmes liés à une éventuelle modification extensive de la TUA n'avaient guère été abordés². Cependant cet avis doit être nuancé, notamment du fait que nous avons trouvé dans un texte de Gilles Gaston Granger de 1967 : *Pensée formelle et sciences de l'homme*, des éléments indubitables qui montrent que le grand épistémologue français avait eu une intuition très nette et très précise de ce qui allait devenir la Théorie de la Perspective. À notre connaissance il s'agit de la seule anticipation aussi claire de ladite théorie³ :

« Mais on peut se demander si l'évaluation des perspectives ne serait pas plus adéquatement réalisée en considérant d'une certaine manière comme liées attente et satisfaction ; de telle sorte que, par exemple, *au voisinage de la certitude – et de l'impossibilité – les estimations de satisfactions soient majorées ou minorées selon une loi déterminée*. Dans une région assez éloignée de ces zones critiques tout se passerait comme si attentes et satisfaction étaient indépendantes ; *il faudrait donc concevoir une sorte de schéma relativiste introduisant une «courbure» de la variété satisfaction - attente*, en un sens analogue à celui de l'espace-temps einsteinien. Un tel raffinement de l'axiomatique n'a, croyons-nous jamais été tenté »⁴.

¹ Le titre complet de l'article est : « Prospect Theory : An Analysis of Decision under Risk ». *Econometrica*, 47:2, 263-91. Comme le remarque Kahneman dans la préface de *Choices, Values, and Frames* (2002), p X, la théorie de la perspective était en fait presque totalement achevée dès 1975, sous le titre de « Value Theory », cependant l'élaboration de la phase d'édition prit encore trois années. Le nom de « Value Theory » était sans aucun doute dû au rôle majeur dévolu à la fonction de valeur (« Value Function »). Toutes nos citations sont issues de la réédition de l'article dans le recueil *Choices, Values, and Frames*. Nous citerons K/T (1979). La littérature secondaire étant encore une fois plus qu'abondante puisque cet article est l'un des plus cités au monde dans le domaine de l'économie et de la psychologie, nous l'introduisons au fur et à mesure de notre progression analytique. Nous mentionnons cependant comme particulièrement éclairantes du point de vue pédagogique, les approches de Baron (1988), p 362-272 ; Wu/Zhang/Gonzales in Koehler/Harvey (Eds.) (2004), p 399-423 ; Peterson (2009), p 285-295 ; et enfin Gilboa (2009), p 157-159. Une bonne introduction peut être aussi trouvée dans l'article de Levy (1992). L'intégration et l'utilisation de la TP dans le cadre de l'économie fut le fait de Thaler (1980). L'imposant livre de Wakker (2010) demeure toutefois ce qui a été écrit de plus achevé sur le sujet, le seul inconvénient c'est qu'il traite presque exclusivement de la Théorie Cumulative de la Perspective, et que son formalisme est très lourd, c'est un ouvrage fondamental dans le domaine de l'économie.

² Il y avait cependant le livre de Saint-Sernin qui était paru en 1973, et la littérature de langue anglaise était déjà substantielle, comme en témoignent les contributions de Davidson, Suppes, Jeffrey ou Nozick. Mais ces contributions tournaient souvent autour de la portée philosophique de la TUA et de sa falsifiabilité, du bayésianisme, et non pas sur une possible modification de cette dernière à la façon de ce qu'allaient faire K/T.

³ Ceci est d'autant plus pertinent que l'extrait se situe après la présentation de quelques problèmes liés à l'axiomatique de N/M, notamment ceux liés à la représentation des perspectives.

⁴ Granger (1967), p 173-174. C'est nous qui soulignons. Il est important de noter que l'auteur parle de modifier la TUA exactement dans le sens où irons K/T plus tard, c'est ce qui fait la différence de son anticipation avec par exemple celle d'Allais ou d'Edwards.

L'article est divisé en quatre parties d'inégales longueurs et un appendice¹. Il commence par quelques remarques sur le caractère normatif de la TUA. Contrairement à cette dernière théorie, la TP sera une théorie descriptive² qui formalisera elle aussi le concept de choix rationnel en contexte d'incertitude. Nous restons donc dans un cadre assez classique puisqu'il s'agira par exemple de considérer les perspectives comme les alternatives d'une loterie de forme : $(x, p ; y, 1-p)$. Mais il s'agira surtout de présenter une série de contestations empiriques basées sur des questionnaires soigneusement construits : « ...in which preferences systematically violate the axioms of expected utility theory »³. Le but de la théorie sera donc double : une partie critique qui visera en quelque sorte à raffiner les réfutations empiriques de la TUA commencées par Allais, d'autre part, il s'agira pour K/T, de construire une théorie, aussi conservatrice que possible, capable d'intégrer les résultats expérimentaux obtenus jusqu'ici. Pourquoi K/T ont-ils tenu à élaborer une théorie qui reste de par sa forme aussi proche de la TUA que possible ? Pourquoi ne pas changer le paradigme ? Kahneman répond à cette question dans la préface de *Choices, Values, and Frames*⁴ : par souci de simplicité, il fallait que l'ensemble des modifications soit minimal afin de pouvoir continuer à décrire des situations décisionnelles extrêmement simples élaborées sur le type de la loterie.

La partie proprement critique de l'article commence par une présentation sommaire⁵ des caractéristiques principales de la TUA :

- Expectation : l'utilité générale d'une perspective est équivalente à l'utilité attendue de ses résultats, calculée en multipliant les probabilités et les gains :

$$U(x_1, p_1 ; \dots ; x_n, p_n) = p_1 u(x_1) + \dots + p_n u(x_n).$$

¹ Appendice qui présente l'axiomatique sous-tendant la TP.

² Mais elle aura toutefois une portée normative, notamment grâce au concept d'aversion au risque, et une portée parfois même assez inattendue comme en science politique, voir Levy (2003). Ces extensions avaient été anticipées par Kahneman, voir la préface de *Choices, Values, and Frames*, p XI.

³ K/T (1979), p 18.

⁴ P X.

⁵ K/T (1979), p 18.

- Intégration de l'avantage : une perspective donnée $(x_1, p_1; \dots; x_n, p_n)$ est acceptable pour un sujet avec un avoir présent w si et seulement si :

$$U(w + x_1, p_1; \dots; w + x_n, p_n) > u(w).$$

Ce qui implique que le domaine de la fonction d'utilité soit l'état final de la fortune du sujet¹.

- Aversion aux risques due à la concavité de la fonction d'utilité : ($u'' < 0$) . Comme nous le remarquons dans notre analyse du paradoxe d'Allais, l'aversion au risque peut être définie par le fait de préférer un gain certain à toute loterie ayant le même gain comme utilité attendue.

Ici deux remarques s'imposent. Tout d'abord K/T ne présentent pas les axiomes de la TUA proprement dits, ce qui pourrait sembler de prime abord étrange, d'autre part, les auteurs ne présentent pas non plus d'emblée les deux caractéristiques importantes de la TUA que sont l'invariance et la dominance², et qui vont être amenées à jouer un rôle important, l'une lorsque les auteurs vont analyser l'effet de cadrage, l'autre lorsqu'ils présenteront les opérations de la phase d'édition. Après cette présentation, K/T font état de quelques remarques méthodologiques, sur lesquelles nous nous arrêtons brièvement. Il s'agit de s'interroger sur la pertinence de l'utilisation des questionnaires hypothétiques³ qui vont être construits sur le modèle suivant :

A : 50 % de chance de gagner 1000.

B : 450 sûrs.

50 % de ne rien gagner.

Les tests furent présentés à deux groupes d'étudiants israéliens⁴. Il s'agissait pour eux d'imaginer qu'ils étaient réellement en face de la situation proposée, et de dire quel est le

¹ Cette clause avait été remise en question par Markovitz (1952), et ce sera un des aspects fondamentaux de la TP, par l'introduction de la fonction de valeur, de poser que le domaine de la fonction d'utilité soit les gains et les pertes par rapport à un point de référence. Voir Baron (1988), p 267-271 ; tout le troisième chapitre de *Choices, Values, and Frames*, p 143-208, est consacré à l'élaboration du concept de fonction de valeur.

² K/T (1986), p 210.

³ Savage (1954), p 28-29, avait émis des doutes quant à la pertinence de cette méthode qui repose au fond sur l'introspection.

⁴ Il s'agit donc de devises israéliennes, à l'époque le salaire moyen d'une famille était de 3000 \$.

choix qu'ils auraient fait. Cette méthode a des défauts évidents, notamment le fait qu'au fond rien ne garantit l'adéquation entre le choix hypothétique et le choix réel. Cependant, et c'est là un des tours de force de K/T, les questionnaires avaient été élaborés de façon à ce que les sujets puissent répondre le plus naturellement possible. Comme nous allons le voir les options étaient réalistes sans engager des sommes extraordinaires. La méthode n'est certes pas parfaite mais elle contourne le problème de l'artificialité¹ des tests en laboratoire et la nécessaire répétition des expériences. Par ailleurs l'expérience directe ne peut donner qu'une masse d'informations disparates difficiles à mettre en forme. Le questionnaire se situe donc comme à mi-chemin entre les deux méthodes, son utilisation repose sur l'hypothèse que : « People often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reasons to disguise their true preferences »².

La première partie de l'article est intitulée : « Certainty, Probability, and Possibility », elle a pour but de mettre en avant la violation du PCS et de l'axiome de substitution-indépendance en faisant intervenir pour le premier l'effet de la conséquence commune, et l'effet du rapport commun³ pour le second. Il s'agit de remettre en cause le fait que les choix sont indépendants des distributions de probabilités, c'est une idée qui, comme nous l'avons vu, était déjà présente chez Allais, Edwards, Ellsberg et même L/S de façon plus détournée. Cependant la façon dont les perspectives sont construites par KT va leur permettre d'échapper aux critiques d'Amihud⁴ concernant à la fois l'irréalisme des sommes et des probabilités engagées. Le fait que les sujets surévaluent systématiquement les perspectives certaines au

¹ C'est un problème réel, tant du point de vue philosophique que du point de vue de la généralisation statistique. Par ailleurs, le fait de provoquer les résultats, c'est-à-dire les violations de la TUA ajoute une dimension artificielle dans le processus. C'est aussi une critique qui sera adressée au HB, notamment par Gigerenzer (1991) : les biais sont provoqués par la forme des questionnaires eux-mêmes, et ils disparaissent quand on modifie les données du problème, notamment lorsque ce sont les données probabilistes qui sont modifiées.

² K/T (1979), p 19.

³ Concernant ces concepts, voir notre partie sur Allais ; mais aussi Machina (1987), p 127-130 ; Willinger (1990), p 12-15 ; Wu/Zhang/Gonzalez in Koehler/Harvey (Eds.) (2004), p 403-407 ; Sugden in Barberà/Hammond/Seidl (Eds.) (2004), p 694-699.

⁴ Voir Amihud in Allais/Hagen (Eds.) (1979), p149-160 ; et Machina (1987), p 129. L'article de Amihud est sorti la même année que la TP, les auteurs n'ont donc pas pu le compiler, cependant, il semble évident qu'ils étaient parfaitement au courant des problèmes liés au réalisme dans la présentation qu'avait fait Allais de son paradoxe. C'est un fait intéressant du point de vue de l'épistémologie et de la falsifiabilité de la TP, comme nous allons le voir, la structure est exactement la même que dans le paradoxe, pourtant le réalisme des options ajoute beaucoup à la pertinence et nous fait sortir des cas « isolés » présentés par Allais. Avec la TP, les situations de ce type deviennent possibles et il ne suffit plus d'évacuer les violations de la TUA d'un revers de la main en arguant de l'irréalisme. Il nous semble que ce point philosophique concernant le réalisme de la TP n'a pas été suffisamment mis en avant.

détriment des perspectives aléatoires, même si elles ont une utilité attendue plus importante sera baptisé par K/T « effet de la certitude » (« certainty effect »¹). Les choix sont les suivants :

P1 : Choisir entre,

A :	2500, .33 2400, .66 0, .01	B :	2400, 1
N = 72	[18]		[82]*

P2 : Choisir entre,

C :	2500, .33 0, .67	D :	2400, .34 0, .66
N = 72	[83]*		[17] ²

Si nous nous fions à l'aspect prédictif du principe d'indépendance nous devrions observer les choix suivants : si B est préféré à A alors D doit être préféré à C du fait de l'indépendance des préférences et des probabilités. Or, comme dans le paradoxe d'Allais et le paradoxe d'Ellsberg il n'en est rien. Les préférences des sujets se portent en premier lieu vers l'option sûre, préférant ne pas risquer le (0, .01) qui pourrait tout leur faire perdre, par contre dès que nous nous éloignons de la certitude, comme l'avaient d'ailleurs établis L/S³, c'est le gain le plus élevé qui va être choisi.

D'autre part si nous analysons les choses de plus près, nous voyons que la préférence ainsi obtenue est incompatible avec l'élaboration d'une fonction d'utilité de type N/M, puisque dans le **P1** nous avons l'inégalité suivante :

$$u(2400) > .33u(2500) + .66u(2400)$$

¹ Tversky/Fox (1995), p 97, formalisent très clairement cet effet dans le cadre de la TP et de la TCP.

² N correspond au nombre de participants et le chiffre entre crochets correspond au pourcentage de sujets qui ont choisi l'option.

³ Rappelons-nous que ces auteurs avaient montré que le stimulus qui va orienter le choix peut être la probabilité ou les gains, mais que ce sont surtout les hautes probabilités qui sont considérées en premier lieu, elles ont un impact plus fort sur la représentation subjective : elles jouent le rôle d'une forme de garantie par rapport au gain, voir L/S (1971).

Si nous soustrayons $.66u(2400)$ des deux options de **P1** :

$$u(2400) - .66u(2400) > .33u(2500)$$

Nous obtenons donc pour les préférences exprimées par les sujets :

$$.34u(2400) > .33u(2500)$$

Or l'inégalité construite à partir des préférences des sujets sur **P2** (qui est obtenue en enlevant $.66u(2400)$ à **P1**) est :

$$.34u(2400) < .33u(2500)$$

C'est-à-dire l'inégalité inverse. Nous avons donc affaire, de façon structurelle à un *renversement des préférences*. La subtilité est qu'ici la *conséquence commune* $.66u(2400)$ n'apparaît pas de façon explicite dans l'option certaine, pourtant elle y est incluse de façon nécessaire. Tout se passe exactement comme dans le paradoxe d'Allais¹, la suppression de la conséquence commune entraîne un renversement des préférences, d'où le nom d'« effet de la conséquence commune ». Or c'est une violation nette du principe d'indépendance et du PCS.

Il convient de remarquer ici que l'impact de la *certitude* sur la capacité représentative du sujet est surévaluée par rapport à la *possibilité*, et ce pour la raison suivante : bien qu'il y ait solution de continuité du point de vue de la théorie des probabilités entre le possible et le certain², il n'en va pas de même pour les sujets, le passage du certain au probable, est conçu comme un véritablement *changement d'échelle* et de catégorie. Ces deux catégories affectent la capacité représentative du sujet, l'une en faisant intervenir le concept de certitude, l'autre en faisant intervenir le concept de risque³. Le changement de catégorie produit une : « reduction in desirability when it alters the character of the prospects from a sure gain to a probable one »⁴. Il conviendra de s'interroger dans notre dernière partie sur le rôle que peuvent jouer

¹ Voir notre analyse du paradoxe, c'est le même mécanisme qui cause aussi le paradoxe d'Ellsberg.

² Voir aussi Tversky/Fox (1995), p 97, pour la formalisation de cette idée en TP et en TCP.

³ L'intuition sous-jacente à ce phénomène pourrait être formulée avec le dicton « Un tiens vaut mieux que deux tu l'auras ».

⁴ K/T (1979), p 20-21.

les concepts de certitude, de possibilité et de probabilité dans le cadre élargi du SRD, et de se demander aussi dans quelle mesure le probable peut être la mesure du possible¹ ?

K/T passent ensuite à la description de l'effet du *rapport commun*, qui sera mis en évidence grâce à l'utilisation de loteries simples à deux résultats. Les choix sont les suivants :

P3 : Choisir entre

A :	(4000, .80)	B :	(3000)
N = 95	[20]		[80]*

P4 : Choisir entre

C :	(4000, .20)	D :	(3000, .25)
N = 95	[65]*		[35]

La situation est construite de telle façon que lorsque nous passons de **P3** à **P4** les probabilités sont divisées par 4 (ce qui explique le nom de « common ratio effect »²), ce qui devrait garantir que si B est préféré à A alors D doit être préféré à C, puisque tel que l'indique l'axiome d'indépendance-substitution³ : « The substitution axiom of utility theory asserts that if B is preferred to A, then any (probability) mixture (B, p) , must be preferred to the mixture (A, p) »⁴. Comme dans l'exemple précédent, l'indépendance des gains par rapport aux probabilités n'est pas respectée, en raison de l'effet certitude qui ajoute une dimension psychologique. Comme nous le verrons plus tard, ce phénomène sera expliqué par le fonctionnement du coefficient de pondération sur les probabilités. C'est d'ailleurs en partie la perception biaisée des probabilités par les sujets qui explique le phénomène : réduire la probabilité de gagner de 1 à .25 a plus d'impact que la réduire de .8 à .2. C'est le passage de la certitude à la possibilité qui encore une fois est le ressort dans l'élaboration de la préférence. Comme le remarquent les auteurs, l'effet du rapport commun fonctionne aussi avec des

¹ Pour une très bonne analyse de ces concepts voir Granger (1995), p 129-165.

² Pour une formalisation plus précise et une discussion plus détaillée de l'effet de rapport commun voir notre partie sur Allais ; voir aussi Prelec (2000), p 71 ; et Wu/Zhang/Gonzalez in Koehler/Harvey (Eds.) (2004), p 405.

³ K/T font ici référence en fait à l'axiome de substitution de Samuelson (1952), ils suivent en cela le même plan qu'Allais dans son article, voir notre partie sur cet auteur. Rappelons pour mémoire que l'axiome dans la formulation de Samuelson, de façon originelle, a la forme suivante : $(\forall u)(\forall v)(\forall w) \in U[(u \succ v) \Rightarrow [(u/2) + (1/2)w \sim (v/2) + (1/2)w]]$, forme jugée trop limitative qui fut ensuite généralisée comme suit : $(\forall u)(\forall v)(\forall w) \in U[(\forall \alpha \in [0,1])[(u \succ v) \Rightarrow (u\alpha + (1-\alpha)w \succ v\alpha + (1-\alpha)w)]]$.

⁴ K/T (1979), p 21.

résultats non-monétaires. Il fonctionne avec n'importe quel type de résultat du moment que les probabilités de les atteindre soient divisées selon le même rapport.

L'axiome de substitution subit encore un autre type de violation, qui cette fois est à mettre en rapport avec les résultats des expériences de L/S concernant les stimuli – c'est-à-dire ici l'importance relative des gains par rapport aux probabilités et réciproquement – qui interviennent et qui influent sur le choix. Encore une fois nous nous situons au niveau représentationnel¹. Soit le problème suivant² :

P7 : Choisir entre

A :	(6000, .45)	B :	(3000, .90)
N = 66	[14]		[86]*

P8 : Choisir entre

C :	(6000, .001)	D :	(3000, .002)
N = 66	[73]*		[27]

Encore une fois, les préférences des sujets reflètent leurs façons de se représenter les probabilités : le passage d'une probabilité représentable à une probabilité si minuscule que les sujets ne voient plus très bien à quoi elle correspond réellement³ entraîne encore une fois un renversement des préférences. Pourtant ici aussi nous passons de **P7** à **P8** en divisant les probabilités par le même facteur⁴ ce qui ne devrait donc pas induire de renversement. Lorsque les probabilités cessent de pouvoir être représentées de façon intuitive, c'est-à-dire qu'elles se situent en deçà d'un certain seuil⁵, ce sont les gains seuls qui sont pris en compte dans la construction des préférences. Comme le remarquent KT : « In this situation where winning is possible but not probable, most people choose the prospect that offers the larger gain »⁶.

¹ Tous les problèmes que K/T ont soumis à leurs sujets subissent les effets d'au moins deux filtrages représentationnels qui seront formalisés comme le coefficient de pondération et la fonction de valeur. C'est là que réside l'originalité réelle de la TP, car les contestations en tant que telles, et même les effets de cadrages, étaient déjà connues. Le problème est d'ailleurs évoqué par Savage (1954).

² Cet exemple doit donc être lu à la lueur des analyses de L/S sur les paris P et les paris \$, voir L/S (1971), p 53.

³ Il serait intéressant de se demander à quel moment se situe précisément le passage du représentable au non représentable.

⁴ En l'occurrence 45000 ici...

⁵ Nous verrons aussi que quand elles se situent au-delà d'un certain seuil elles sont assimilées à la certitude.

⁶ K/T (1979), p 22. Remarquons que c'est le « jeu » entre le possible et le probable qui est le ressort du phénomène. Il faudra donc analyser cette relation de plus près dans notre troisième chapitre.

Allais (1951) et Edwards (1954) avaient évoqué la question de la perception et de l'évaluation des perspectives avec une utilité attendue négative. Allais parlait surtout d'aversion pour le risque et Edwards d'une pondération probabiliste différente pour les gains et les pertes. K/T vont reprendre le problème d'une façon tout à fait originale afin de faire ressortir deux phénomènes très importants, tout d'abord l'*effet miroir* (« reflection effect ») qui permettra lorsqu'il sera mis en relation avec le concept de fonction de valeur de mettre en avant un autre phénomène central : le fait que du point de vue du système de référence subjectif, les pertes pèsent plus lourd que les gains¹.

K/T se posent la question suivante : que se passe-t-il si nous renversons le signe des gains dans les problèmes que nous venons d'aborder, c'est-à-dire si nous passons du choix entre plusieurs perspectives positives à un choix entre plusieurs perspectives négatives ? La réponse à cette question est donnée dans le tableau ci-dessous² :

Perspectives positives

$$\begin{array}{l} \mathbf{P3} : \quad (4000, .80) < (3000) \\ N = 95 \quad [20] \quad [80]^* \end{array}$$

$$\begin{array}{l} \mathbf{P4} : \quad (4000, .20) > (3000, .25) \\ N = 95 \quad [65]^* \quad [35] \end{array}$$

$$\begin{array}{l} \mathbf{P7} : \quad (3000, .90) > (6000, .45) \\ N = 66 \quad [86]^* \quad [14] \end{array}$$

$$\begin{array}{l} \mathbf{P8} : \quad (3000, .002) < (6000, .001) \\ N = 66 \quad [27] \quad [73]^* \end{array}$$

Perspectives négatives

$$\begin{array}{l} \mathbf{P3'} : \quad (- 4000, .80) > (- 3000) \\ N = 95 \quad [92]^* \quad [8] \end{array}$$

$$\begin{array}{l} \mathbf{P4'} : \quad (- 4000, .20) < (- 3000, .25) \\ N = 95 \quad [42] \quad [58]^* \end{array}$$

$$\begin{array}{l} \mathbf{P7'} : \quad (- 3000, .90) < (- 6000, .45) \\ N = 66 \quad [8] \quad [92]^* \end{array}$$

$$\begin{array}{l} \mathbf{P8'} : \quad (- 3000, .002) > (- 6000, .001) \\ N = 66 \quad [70]^* \quad [30] \end{array}$$

¹ Allais avait eu l'intuition très nette de ce phénomène qui s'exprime dans le paradoxe à travers l'aversion des sujets au risque. Par ailleurs, comme le remarque Kahneman dans la préface de *Choices, Values, and Frames*, p XIII, ce fait est une des découvertes les plus importantes de la TP. Ce en quoi nous lui donnons entièrement raison puisque nous pensons pouvoir donner une explication d'un certain type d'akrasie à partir de l'exploitation de ce phénomène et du phénomène d'aversion au risque.

² Il est tiré de l'article de 1979, p 22.

Le premier élément qui émerge ici, c'est le fait que les préférences, qui dans le domaine des gains révélaient une *aversion au risque*, s'inversent de façon systématique avec le signe des résultats et mettent en évidence un phénomène de *recherche du risque* dans le domaine des pertes. Par exemple si nous prenons **P3**, nous observons que c'est la perspective sûre qui est préférée mais que quand nous passons vers une perspective avec une utilité négative, c'est le pari qui est préféré, pari qui risque de faire perdre plus au sujet que la perte sèche de 3000¹. Or ce phénomène est incompatible avec l'hypothèse de maximisation de la TUA. L'effet miroir met donc au jour quatre manières qu'ont les sujets d'envisager le risque, qui peuvent être résumées dans ce tableau² :

	<i>Petites probabilités</i>	<i>Moyennes/grandes probabilités</i>
Gains	Recherche du risque	Aversion au risque
Pertes	Aversion au risque	Recherche du risque

Il y a donc une *asymétrie assez nette entre les pertes et les gains*³, asymétrie que viendra formaliser la fonction de valeur. Par ailleurs si nous partons du principe que les pertes pèsent plus lourd que les gains, comment se fait-il que les sujets choisissent de façon systématique les options qui risquent de leur faire perdre la somme la plus grande ? Comme le remarquent K/T, c'est encore une fois la surestimation de la certitude qui est en cause ici. Dans le cadre de paris avec utilité attendue négative, la perspective d'une perte certaine pousse les sujets à se réfugier dans une option dont le résultat incertain risque de leur faire perdre plus encore. Ce qui se passe ici, notamment dans les exemple **P3** et **P3'**, c'est que le .20 de ne rien perdre est à son tour surévaluer face à la possibilité de .80 de perdre 4000. Nous pouvons ici

¹ En l'occurrence le sujet risque ici de perdre 3200 dollars. Ce phénomène avait été documenté par Markovitz (1952) et des expériences avaient commencé à être menées par Williams (1966), expériences dans lesquelles les sujets étaient par exemple indifférents entre : (100, .65 ; -100, .35) et (0, sûr), ce qui reflète une aversion pour le risque puisque l'utilité attendue du pari est supérieure (+35). Le même auteur avait mis au jour une indifférence dans le domaine de l'utilité négative entre (-200, .80) et (-100, sûrs), ce qui dénote une recherche du risque puisque le pari a une utilité attendue négative plus importante que l'option sûre (-160).

² Inspiré par Wakker/Tversky (1995), p 1256.

³ Gilboa (2009) pense que c'est là la découverte la plus importante de la TP, nous aurons en effet l'occasion de montrer qu'elle a une portée beaucoup plus générale que le laisse entendre l'article de 1979. La crainte de perdre – et la surestimation des pertes par rapport aux gains –, dans sa généralité, est un élément qui structure nos existences. C'est l'exact opposé du stoïcisme et du bouddhisme zen.

faire l'hypothèse que ce n'est pas .20 de ne rien perdre que les sujets envisagent ici, c'est autre chose, ils envisagent seulement la possibilité de ne rien perdre et non la probabilité telle qu'elle est établie¹. Il s'opère ici probablement une pondération d'un genre particulier, liée au concept de perte, qu'il faudrait formaliser plus rigoureusement, puisque le même effet produit en fait deux phénomènes opposés : « The same psychological principle – the overweighting of certainty – favors risk aversion in the domaine of gains and risk seeking in the domaine of losses »². Il s'agit là d'un phénomène central pour toute la théorie de la décision, et il conviendra de l'analyser plus à fond dans notre troisième chapitre en vue de l'intégrer à nos propres hypothèses, car cet effet se donne à voir et à penser dans d'autres contextes que celui de la TP. *Il capture un élément important dans l'explication du système de référence décisionnel, à savoir la façon dont les concepts de pertes et de risques influencent le fonctionnement de ce système.* C'est une réflexion qui doit être menée in extenso, car elle a une grande portée philosophique : dans quelle mesure pouvons-nous dire que nos décisions – quotidiennes ou pas – sont influencées, motivées par les concepts de risque et de pertes ?

Un dernier point que soulèvent K/T concerne le fait que l'effet miroir élimine l'incertitude ou la variabilité des perspectives comme explication de l'effet certitude. Il s'agirait ici de penser que les sujets préfèrent les perspectives qui ont 1) une variance minimale, 2) une utilité attendue élevée. C'est la thèse soutenue par Allais dans son article de 1953. Par exemple dans **P3**, le fait que 3000 n'ait aucune variance expliquerait que le choix des sujets ne se porte pas sur le pari qui a pourtant une utilité attendue plus élevée, cela expliquerait aussi que lorsque les probabilités sont divisées par quatre dans **P4** les sujets préfèrent le pari : dans ce cas-là, la variance des deux perspectives serait trop minime pour compenser la différence entre les utilités. Or si c'était le cas lorsque l'on passe à une utilité négative, c'est l'option qui a le moins de variance qui devrait être choisie, c'est-à-dire l'option sûre, d'autant plus que c'est elle qui a l'utilité attendue la plus élevée. Mais les sujets choisissent le pari. Il apparaît donc que : « The certainty increases the aversiveness of losses as

¹ Ce qui nous renvoie encore une fois aux représentations subjectives du possible et du probable qui devront être analysées de plus près. Voir à ce sujet les intéressantes remarques de Prelec (2000), p 79.

² K/T (1979), p 23. Le fait que la surestimation de la certitude puisse rendre compte de deux effets aussi opposés pourrait rendre sa portée explicative suspecte, nous serions un peu en face du même cas de figure qu'avec la critique de la portée explicative de l'heuristique de représentativité pour le sophisme du joueur et le sophisme de la bonne main, voir Gigerenzer/Brighton (2009), p 18.

well as the desirability of gains »¹.

La prochaine étape de la critique² vise à mettre en avant deux éléments : la violation de la propriété d'invariance³, et – de façon plus implicite – le fonctionnement de l'effet de cadrage. Il appert que les sujets éliminent parfois spontanément les composantes des options que les alternatives partagent, ce qui peut les mener à certaines incohérences. Pour bien cerner ce phénomène il faut se rapporter aux données de **P4**, où le sujet avait le choix entre : (4000, .20) et (3000, .25).

Dans le problème 10 (**P10**) Kahneman et Tversky demandent aux sujets de considérer un jeu à deux niveaux dans lequel il y a d'abord ou .75 de terminer sans rien ou .25 d'accéder à la deuxième phase du jeu, auquel cas les sujets devront choisir entre (4000, .80) et (3000). Le choix concernant la deuxième étape doit être fait *avant* que le jeu ne commence. Lorsque nous calculons les probabilités nous voyons que le jeu à deux étages est en fait strictement équivalent à **P4**⁴. La seule chose qui change c'est le moment où les sujets doivent opérer le choix :

« In the standard form [**P4**] the decision maker faces a choice between two risky prospects, whereas in the sequential form [**P10**] he faces a choice between a risky and a riskless prospect »⁵.

Or les préférences des sujets se trouvent renversées quand l'on passe d'un problème à l'autre, pourquoi ? Parce que le premier moment du jeu est ignoré dans la construction des préférences, moment dans lequel les « outcomes are shared by the prospects ». Le jeu est donc simplifié à un choix entre 3000 sûr et .80 de gagner 4000, le .75 de ne pas passer au second niveau est ignoré⁶.

¹ K/T, *ibid*.

² Nous laissons volontairement de côté le problème de l'assurance probabiliste, qui de l'aveu des auteurs est construit de façon un peu artificielle. Cet exemple vise à montrer qu'une assurance probabiliste réduisant le risque des pertes de p à $p/2$ devrait, selon la TUA, être préférée à une assurance normale réduisant le risque des pertes de $p/2$ à 0, or les sujets manifestent des préférences nettes pour l'assurance standard dont l'utilité attendue est plus faible, étant donné le rapport du montant de l'assurance aux probabilités. Pour plus de précisions, voir K/T (1984), p 9.

³ Propriété qui repose sur le fait que dans la TUA, la construction des préférences repose sur la seule considération des probabilités des états finaux.

⁴ Voir K/T (1979), p 26 pour les deux types de formulations (standard et séquentielle).

⁵ K/T, *ibid*.

⁶ K/T avaient déjà mis en avant ce phénomène dans leur présentation de l'heuristique de représentativité. Par exemple dans le fameux « problème des taxis », la taille de l'échantillon est complètement ignorée, de la même façon que dans le problème de

Cet exemple nous montre deux choses : tout d'abord que les choix des sujets ne sont pas seulement guidés par des considérations probabilistes, comme l'avaient déjà noté Allais, Ellsberg, Lichtenstein et Slovic. Ce fait met encore une fois en défaut la TUA. D'autre part, le principe d'invariance, qui stipule que les choix des sujets ne doivent pas être affectés par la façon dont les options sont présentées, est lui aussi violé.

Parallèlement, la façon dont les rétributions sont présentées altère aussi les préférences. C'est ce que montre les problèmes **11** et **12** : en plus de leur avoir présent on donne aux sujets 1000 \$ et on leur demande de choisir entre (1000, .50) ou B (500), dans un deuxième temps, en plus de ce qu'ils ont déjà on leur donne 2000 et on leur demande de choisir entre C (-1000, .50) et D (-500). La majorité des sujets choisissent B dans le premier cas et C dans le second or : $A = (2000, .50 ; 1000, .50) = C$ et $B = (1500) = D$. Les deux options sont strictement équivalentes mais les sujets éliminent les bonis de 1000 et 2000 et se focalisent sur les choix offerts. Encore une fois, remarquons que c'est le premier moment du processus qui est éliminé, et cela veut dire qu'il y a toute une strate d'information qui n'est pas prise en compte, c'est un peu le même phénomène que nous avons observé dans le cadre des heuristiques et des biais et nous pouvons d'ores et déjà postuler que cette information tronquée aura des conséquences diverses dans le domaine de la prise de décision. Par ailleurs, ces données sont incompatibles avec la TUA à un autre niveau : ce ne sont pas seulement les états finaux de l'avoir du sujet qui sont pris en compte, mais la variation par rapport à un point de référence. L'utilité serait donc liée au changement de fortune, comme le remarquent les auteurs, dans la TUA :

« The same utility is assigned to a wealth of 100000\$, regardless of whether it was reached from a prior wealth of 95000\$ or 105000\$. Consequently, the choice between a total wealth of 100000\$ and even chance to own 95000\$ or 105000\$ should be independent of whether one currently owns the smaller or the larger of these two amounts »¹.

L'élimination des composantes communes aux options qui pousse le sujet à évaluer de façons différentes des perspectives identiques nous conduit au concept de « framing effect »

l'hôpital avec la variation du nombre de naissances. Ce qui met en avant ces phénomènes c'est qu'il y a une strate de l'information qui est totalement négligée, or cette information est déterminante. Nous pensons qu'ici, c'est le même phénomène qui intervient et qui produit le renversement des préférences : la première étape du jeu est totalement ignorée.

¹ K/T (1979), p 28.

lui-même lié au concept de « représentation ». L'effet de cadrage¹ – ou effet du cadre de référence – peut être entendu de deux façons différentes : il peut s'agir de la façon dont le sujet construit sa représentation des alternatives, ou il peut s'agir de la façon dont les alternatives lui sont présentées de l'extérieur.

Ces deux versions du même effet ont des conséquences assez importantes dans le domaine de la prise de décision, la première de ces versions peut être mise en relation avec l'heuristique d'ancrage. La seconde version pourrait nous conduire à une réflexion sur le problème de la manipulation. L'effet de cadrage est en tout cas à verser à la problématique générale de la représentation qui comme nous avons commencé de le voir est centrale en théorie de la décision. C'est d'ailleurs probablement pour arriver à décrire le fonctionnement de certains éléments représentatifs que K/T en sont venus à formaliser le coefficient de pondération et la fonction de valeur.

b – L'élaboration de la théorie.

+ La phase d'édition et l'introduction du coefficient de pondération et de la fonction de valeur.

Une des originalités de la TP c'est qu'à la différence de la TUA, dont elle représente l'alternative² la plus aboutie, elle propose que les décisions se déroulent en deux phases distinctes³, qui permettent chacune d'expliquer le comportement des sujets vis-à-vis des options qui leurs sont présentées. Le premier moment de la prise de décision est celui de la phase d'édition qui consiste en une « preliminary analysis of the offered prospects, which often yields a simpler representation of the offered prospects »⁴. Comme nous le remarquons plus haut, nous pensons que cette phase d'édition doit être mise en relation avec le concept d'heuristique tel que nous l'avons défini dans la partie précédente. En effet ici aussi l'action de

¹ Voir K/T (1986) ; Manktelow (2012), p 205-206.

² Alternative en ce sens qu'elle va rester très proche de la TUA en ne prenant en compte que les perspectives avec des résultats monétaires et des probabilités établies de façon objective.

³ Il y a probablement une phase d'édition dans la TUA, mais elle n'a fait l'objet d'aucune thématisation explicite de la part de Ramsey, N/M ou Savage. C'est à partir de ces deux phases – auxquelles nous rajouterons une opération de visée – que nous allons construire notre concept de système de référence décisionnel, voir notre schéma, p 219.

⁴ K/T (1979), p 28. Il faut remarquer que cette phase est celle qui a reçu le moins d'attention de la part des chercheurs.

la faculté représentative¹ du sujet va simplifier l'information qui lui est communiquée en faisant intervenir – de façon presque automatique et à la limite de la conscience² – une série d'opérations de types heuristiques dont le but sera de :

« organize and reformulate the options so as to simplify subsequent evaluation and choices. Editing consists of the application of several operations that transforms the outcomes and probabilities associated with the offered prospects »³.

Encore une fois, et comme nous constatons depuis Allais, Edwards et L/S, les sujets ne traitent pas l'information, probabiliste ou ayant rapport aux gains possibles, telle qu'elle leur est transmise. Cette information passe par un *filtre représentatif*, filtre qu'avait commencé à élaborer la théorie des HB, qui va jouer un rôle crucial puisqu'il a pour vocation de transformer et de simplifier les options, c'est à dire tout simplement de les changer. La question qui émerge ici, et qui était déjà présente dans les HB est la suivante : est-ce que la perspective qui va faire l'objet du choix des sujets est la même que celle qui leur a été présentée ? La distorsion induite par la phase d'édition ne rend-elle pas les options et les alternatives autres qu'elles ne sont en réalité ? Ce qui voudrait dire que *les sujets ne décident pas à proprement parler à propos des alternatives qui leur sont présentées, mais des alternatives qu'ils ont construites*.

1) La première de ces opérations⁴ est le codage (« coding »), elle transforme la perception des rétributions en gains ou en pertes par rapport à un point de référence

¹ Nous entendons ici ce concept dans un sens assez large. Ce qu'il faut se demander – ce serait un sujet de recherches à part entière et intéressant – c'est quelles sont précisément les similitudes et les différences entre le fonctionnement de la faculté représentative dans le contexte des heuristiques et des biais, par rapport au contexte de la phase d'édition de la TP ? Autrement dit, peut-on tout simplement assimiler les opérations de la phase d'édition à des heuristiques ? C'est l'hypothèse que nous faisons, mais elle pourrait être plus approfondie et plus travaillée. C'est un sujet qui n'a guère été abordé, la phase d'édition étant très peu analysée ou mentionnée, voir par exemple pour une utilisation de cette partie de la TP Goldstein/Einhorn (1987) ; Birnbaum (2002).

² K/T ne développent le sujet qu'allusivement, en disant que les opérations en question sont mobilisées, à chaque fois que c'est possible, mais la façon dont les opérations sont mobilisées reste obscure : est-on dans un processus semi-automatique ? Probablement que oui dans la mesure où l'habitude l'aura rendu tel, comme pour les heuristiques. Nous ne pensons pas que l'on puisse dire à proprement parler que le sujet choisit de façon totalement réfléchie d'appliquer telle ou telle option.

³ K/T (1979), p 28.

⁴ Ces opérations seront intégrées plus loin dans notre concept de système de référence décisionnel (SRD), dont une des fonctions consiste à éditer les options d'une façon assez similaire à la TP. Cependant comme dans l'existence les perspectives n'arrivent pas toujours avec des mesures probabilistes il faudra aussi arriver à comprendre comment le SRD édite les perspectives à caractère non probabilistes.

donné. Le codage peut aussi faire se déplacer le point de référence par rapport auquel les perspectives sont évaluées en fonction de la manière dont le problème est présenté ainsi qu'en fonction des souhaits du décideur¹. C'est une opération que l'on pourrait rapprocher de l'heuristique d'ancrage et d'ajustement puisque le codage peut être influencé par la façon dont l'alternative est formulée².

2) La combinaison (« combination ») simplifie les perspectives en additionnant les probabilités quand celles-ci ont les mêmes rétributions. Par exemple (200, .25 ; 200, .25) est transformé en (200, .50).

3) La ségrégation (« segregation ») consiste à séparer les composantes risquées des perspectives des composantes sans risques : (300, .80 ; 200, .20) est donc évalué comme un gain sûr de (200) et une perspective risquée de (100 ; .80)³. Il convient ici de se demander dans quelle mesure la perspective risquée est prise en compte par le sujet et comment elle influence la décision. K/T ne le disent pas⁴.

Comme le font remarquer les auteurs : « The preceding operations are applied to each prospect separately. The following operation is applied to a set of two or more prospects »⁵. Nous avons donc ici à faire à une édition commune des perspectives et elle est très importante.

4) Le mécanisme de l'annulation (« cancellation ») est ce qui explique par exemple que dans le jeu à deux étages les sujets ignorent la première phase. Les composantes communes, qu'elles soient de l'ordre des probabilités ou des rétributions, sont éliminées systématiquement. Cela est important dans le sens où dans un processus décisionnel il va y avoir un manque d'information qui va venir biaiser le jugement ou la décision. Comme nous le remarquons plus haut, nous pensons que ce mécanisme est à rapprocher du biais d'ignorance de la taille l'échantillonnage et de l'ignorance des probabilités a priori, biais qui sont liés à l'heuristique de

¹K/T (1979), p 28-29.

² Voir Tversky/Kahneman (1974), p 14.

³K/T (1979), p 29.

⁴ L'hypothèse pourrait être que la perspective risquée est sous-évaluée ou ignorée.

⁵ Ibid.

représentativité¹.

5) Deux autres mécanismes sont à l'œuvre : la simplification (« simplification ») qui va permettre d'arrondir la forme des perspectives ; par exemple (101, .49) devient (100, .50)². Enfin la détection de la dominance (« detection of dominance ») permet de passer en revue les perspectives offertes et d'éliminer celles qui semblent être dominées.

Ces opérations sont « utilisées » par le sujet dès que cela est possible, dans le but de simplifier la représentation des options, elles peuvent intervenir de façon combinée. Parfois le fonctionnement d'une opération peut inhiber le comportement d'une ou plusieurs autres, ou le déclencher. Les auteurs remarquent que ce sujet n'a pas été traité de façon extensive. Ce qu'il faut retenir c'est que la forme de la perspective éditée dépendra de la séquence d'opérations qui auront été pratiquées par le sujet, forme qui peut donc être éminemment variable en fonction des différences subjectives.

K/T vont maintenant introduire les éléments qui font l'originalité de la théorie de la perspective³, à savoir le *coefficient de pondération sur les probabilités* et la *fonction de valeur à partir de laquelle les gains et les pertes sont évalués*. Encore une fois, il convient de remarquer que ces deux éléments sont à relier au concept de représentation, et c'est dans ce sens-là que nous allons les utiliser dans notre troisième chapitre. C'est grâce à ces deux éléments que les auteurs vont pouvoir subsumer et expliquer les anomalies qui hantaient la TUA depuis Allais. Par ailleurs, l'introduction de ces coefficients va permettre de rester le plus proche possible de la TUA.

Le coefficient de pondération (« weighting function »)⁴ associe à chaque mesure probabiliste un coefficient décisionnel $\pi(p)$ qui reflète la façon dont le sujet est affecté par la

¹ Ici nous sommes confrontés à un problème : comment rattacher l'opération d'annulation à l'heuristique en question ? Car ce que l'annulation partage avec les biais mentionnés plus haut c'est l'élimination d'une strate informative, et non pas des éléments liés à la représentativité en tant que telle. Il faudrait donc pousser plus loin les recherches afin de voir si la structure commune ne se trouverait pas encore plus haut dans la chaîne causale.

² Ce mécanisme implique aussi l'élimination des « extremely unlikely outcomes ».

³ Voir, entre autres, Levy (2002), p 174-184 Gilboa (2009), p 154-159 ; Wakker (2010), p 169 et 239. L'introduction de ces deux éléments a donné lieu à un bon nombre d'études qui sont présentées dans le volume *Choices, Values, and Frames*, aux pages 67 à 142 pour le coefficient de pondération, et aux pages 143 à 208 pour la fonction de valeur.

⁴ K/T parlent aussi d'échelle (« scale ») ou de fonction (« function »).

valeur générale de la perspective. Ce coefficient a pour propriété d'être sous-additif¹. Le deuxième élément : la fonction de valeur (« value function »), associe à chaque résultat, à chaque gain un coefficient un nombre $v(x)$ qui reflète la valeur subjective du résultat ; $v(x)$ mesure aussi la déviation que peut représenter le résultat par rapport au point de référence subjectif, c'est-à-dire les gains et les pertes. Nous remarquons donc qu'il s'agit d'*ajouter à la TUA deux éléments* qui ont la capacité d'exprimer des phénomènes liés à la perception et à la représentation des probabilités et des gains par le sujet, et par là même de donner à la TUA une certaine dimension subjective dont elle était dépourvue². Cela va obliger les auteurs à modifier légèrement le formalisme de la TUA afin d'y intégrer ces deux composantes.

Les deux équations que nous allons présenter concernent la formalisation de perspectives simples ayant deux issues non nulles du type : $(x, p; y, q)$. Pour des perspectives régulières³, avec pour les probabilités et les gains les propriétés suivantes : $p + q < 1$, ou $x \geq 0 \geq y$, ou $x \leq 0 \leq y$, l'équation aura la forme suivante :

$$V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y)$$

Il s'agit donc d'une généralisation de la TUA⁴, mais une généralisation qui diffère de la formulation originale en ceci que n'est plus postulée une linéarité des préférences avec les probabilités, puisque celles-ci sont pondérées par un coefficient sous-additif. La généralisation à un nombre quelconque d'issues et au contexte d'ignorance sera le fait de la Théorie Cumulative de la Perspective de 1992⁵.

En ce qui concerne les perspectives strictement négatives ou positives, les choses sont différentes puisque comme nous l'avons vu plus haut lors de la phase d'édition, les perspectives sont décomposées par l'opération de ségrégation en composante risquée⁶ et

¹ C'est-à-dire que $\pi(p) + \pi(1 - p) < 1$. C'est une caractéristique très importante qui aura une grande portée explicative, elle est analysée en profondeur dans Wakker/Tversky (1995), et par Prelec (2002), p 72-74.

² Il s'agit aussi d'ajouter aussi une note de réalisme et de se rapprocher le plus possible de la description de la façon dont les sujets se comportent in situ.

³ Il s'agit de perspectives dans lesquelles on ne trouve aucun gain sûr ni aucune perte sûre.

⁴ La démonstration de l'unicité de la fonction d'utilité est donnée dans l'appendice, elle repose sur l'introduction de trois axiomes : indépendance faible, annulation et solvabilité. Cette dérivation se trouve à la fin de l'article original. Il serait intéressant de pousser plus avant la comparaison avec la dérivation proposée par N/M.

⁵ Voir K/T (1992), p 46-49 ; Prelec (2002), p 68-69 ; Wakker (2010).

⁶ Ce qui peut être perdu ou gagner et qui fait l'objet de la loterie.

composante sans risque¹.

Il va donc falloir formaliser l'effet qu'a cette opération sur la perspective. Perspective qui sera donc évaluée avec pour les gains et les probabilités les propriétés suivantes, qui diffèrent sensiblement du premier cas $p + q = 1$, ou $x > 0 > y$, ou $x < 0 < y$. L'équation aura la forme :

$$V(x, p; y, q) = v(y) + \pi(p)[v(x) - v(y)]$$

Il est notable que dans ce cas-là, le coefficient de pondération soit appliqué seulement à la composante risquée, ce qui formalise de façon adéquate le rapport du sujet au risque et à la possibilité de perdre. Il y a, comme nous le remarquons plus haut une asymétrie assez claire entre les gains et les pertes que vient formaliser cette équation. La valeur d'une perspective strictement négative ou positive est donc équivalente à l'addition de la valeur de la composante sans risque à la différence de valeur entre les issues modifiée par le coefficient de pondération².

Hormis l'introduction de la fonction de valeur et du coefficient de pondération, nous avons donc à faire à la même structure formelle que la TUA, cependant la structure en question va avoir une portée explicative très importante du point de vue des contestations empiriques étudiées jusqu'ici. Le tour de force de K/T c'est de parvenir à formaliser dans les deux équations d'une part la façon dont les sujets se représentent et distordent les probabilités, et d'autre part l'asymétrie fondamentale entre les pertes et les gains, asymétrie qui est liée au système de référence subjectif, c'est-à-dire à la fonction de valeur. Il va s'agir maintenant d'explorer en profondeur les caractéristiques structurelles de ces deux coefficients qui viennent modifier la structure de la TUA, puisque :

« in order to accomodate the effects described in the first part of the paper, we are compelled to assumes that values are attached to changes rather than to final states, and that decision weights do not coincide with stated probability »³

¹ Il s'agit ici du gain minimum certain ou de la perte minimale certaine.

² Comme le remarquent K/T le terme de droite de la deuxième équation est équivalent à : $\pi(p)v(x) + [1 - \pi(p)]v(y)$. La deuxième équation peut donc être réduite à la première si $\pi(p) + [1 - \pi(p)] = 1$, or nous verrons bientôt que cette condition ne peut être réalisée, car le coefficient de pondération est généralement sous additif.

³ K/T (1979), p 31.

+ Le rôle de la fonction de valeur et l'idée d'un système de référence décisionnel ...
subjectif.

Dans le cadre de notre travail, la fonction de valeur de la TP va être amenée à jouer un rôle déterminant, puisqu'elle va nous servir à construire – de façon analogique – notre concept de système de référence décisionnel (SRD). Système qui sera conçu comme le point de référence à partir duquel le sujet fera ses choix. Dans la TP, la fonction de valeur va être reliée à deux concepts clés en théorie de la décision : l'aversion aux pertes et les effets de cadrages¹. Les caractéristiques de cette fonction, qui est un coefficient sur les issues des perspectives, vont être établies en parallèle avec la psychophysique de la perception² :

« Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. (...) The past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point »³.

Le point de référence dont il est question ici sera donc déterminé par les expériences passées du sujet et par ses caractéristiques physiologiques, ce qui nous indique que le point de référence sera aussi singulier que possible⁴. Il s'agit donc ici de transposer le cas de la température dans le cadre de la prise de décision afin de montrer que par exemple, le même montant à gagner pourra être conçu par un riche comme négligeable et pour un pauvre⁵ il représentera la miche de pain qui l'empêchera de mourir de faim. Le montant devra donc être ramené à leurs avoirs présents respectifs et la fonction de valeur aura pour vocation de représenter l'amplitude acceptable des gains et des pertes pour les sujets⁶, en partant du

¹ Voir à ce sujet Tversky/Kahneman (1991).

² Rappelons ici qu'Allais avait utilisé la loi de Weber-Fechner pour construire son concept de seuil psychologique. K/T utilisent le paradigme de la psychophysique de la perception à de multiples reprises, par exemple dans leur article de 1986 sur les effets de cadrages avec l'illusion de Müller-Lyer.

³ K/T (1979), p 32. Pour une analyse de la fonction de valeur dans un contexte un peu élargi voir K/T (1983), p 3 ; et tout le troisième chapitre de *Choices, Values and Frames* ; voir aussi Baron (1988), p 267-271.

⁴ C'est cette singularité qui nous intéressera particulièrement et devra être approfondie. En effet il sera nécessaire de s'interroger sur la façon dont le système de référence décisionnel est construit par l'histoire et l'expérience de chaque individu pour parvenir à penser la singularité de certaines décisions. La question sera alors la suivante : comment l'historique du sujet détermine, en plus de son point de référence, la totalité de certaines décisions. C'est un sujet très épineux.

⁵ Ce phénomène avait déjà été mis en avant par Bernoulli (1737), voir aussi K/T (1983), *ibid.*

⁶ Elle a en fait pour but de représenter une représentation, puisque la façon dont les résultats sont codés comme des gains ou des pertes par rapport à un point de référence est un élément tiré de la représentation subjective.

principe que tout changement de valeur¹ doit être ramené à la position de l'avoir initial.

Du point de vue graphique, comme il est plus aisé de discriminer entre des amplitudes faibles, par exemple entre 5 et 10, qu'entre des amplitudes plus larges, par exemple 105 et 110², la fonction de valeur sera concave au-dessus du point de référence et convexe au-dessous. Du point de vue formel au-dessus du point de référence nous aurons : $v''(x) < 0$, pour $x > 0$; et pour les pertes : $v''(x) > 0$, pour $x < 0$. Cela aura pour but de représenter la décroissance de l'utilité marginale pour les gains et les pertes en fonction de leur magnitude. Comme nous allons le voir il faut encore complexifier la courbe quant à sa forme pour arriver à représenter le fait central que les pertes pèsent plus lourd que les gains. Les éléments définitionnels élaborés jusqu'à maintenant concernaient des perspectives d'où le risque était banni, il s'agissait de gains ou de pertes sûres. Le but de K/T va être à présent de se demander comment la fonction de valeur opère quand nous passons de la certitude au risque³. L'hypothèse des auteurs est qu'elle se comporte de la même façon, ce que va venir appuyer un exemple :

P13 : Choisir entre : (6000, .25) ou (4000, .25 ; 2000, .25)

N = 68 [18] [82]*

P13' : Choisir entre : (-6000, .25) ou (-4000, .25 ; -2000, .25).

N = 64 [70]* [30]

Si nous appliquons la première équation de la partie précédente aux préférences exprimées nous obtenons : $\pi(.25)v(6000) < \pi(.25)[v(4000) + v(2000)]$ pour **P13**, et nous obtenons l'inégalité suivante pour **P13'** : $\pi(.25)v(-6000) > \pi(.25)[v(-4000) + v(-2000)]$. En soustrayant le $\pi(.25)$ des deux termes nous obtenons deux inégalités qui du fait de

¹ Il s'agit donc ici de mesurer l'amplitude du changement subjectif autour d'un point de référence conçu comme le statu quo, et non plus de mesurer la variance autour d'une moyenne comme c'était le cas chez Allais.

² Comme le remarquent les auteurs ceci sera valable dans le domaine des gains comme dans celui des pertes : la différence entre un gain de 100 et un gain de 200 apparaîtra plus grande qu'entre un gain de 1100 et un gain de 1200, idem pour les pertes.

³ Le fait que les auteurs continuent de se mouvoir conceptuellement dans le domaine de la théorie classique de la décision en utilisant des éléments comme le contexte de risque, des gains monétaires, etc., constitue pour nous une limitation essentielle de la TP. Limitation qu'il conviendra de dépasser en partie dans notre troisième chapitre.

l'intervention de l'effet miroir sont en accord avec la définition d'une fonction de valeur concave pour les gains (aversion aux risques) et convexes pour les pertes (recherche du risque)¹.

À la présentation du concept de fonction de valeur doit être rattaché le fait que *les pertes pèsent plus lourd que les gains*. L'expérience de la perte en tant que telle est toujours vécue d'une façon plus intense et plus dramatique que l'expérience d'un gain équivalent. Pour nous il s'agit là d'un fait important. La surévaluation des pertes par rapport aux gains est due au phénomène d'aversion pour le risque.

Comme nous le mentionnions plus haut, nous pensons que cette surévaluation des pertes couplée au concept d'aversion au risque peut expliquer une certaine forme d'akrasie, en particulier des akrasies liées au statu quo et aux préférences habituelles. Nous étudierons ce problème en détail au troisième chapitre. Pour K/T la surévaluation des pertes par rapport aux gains se donne à voir dans le fait que la déception due à la perte d'un montant x n'arrive jamais à être contrebalancée par le gain du même montant². Ce fait explique que les sujets trouvent les paris symétriques du genre : $(x, .50 ; -x, .50)$ inattractifs, et que l'aversion à ce type de paris augmente avec la tailles des enjeux. Ce qui nous indique que du point de vue de la courbe de la fonction de valeur, la pente est plus accentuée pour les pertes que pour les gains, caractéristique qui permet de rendre compte du même coup de la recherche du risque quand les perspectives sont négatives, ou en contexte de pertes répétées³. La fonction de valeur va jouer un rôle central dans l'élaboration de nos hypothèses, comme nous le remarquons dans notre introduction elle va nous servir à construire le Système de Référence Décisionnel à partir duquel les alternatives seront construites et évaluées. Les caractéristiques de cette fonction sont donc les suivantes :

- 1) Elle est définie en fonction de la déviation par rapport à un point de référence, nous parlons donc ici d'une magnitude, ou d'une amplitude, variable et dynamique.

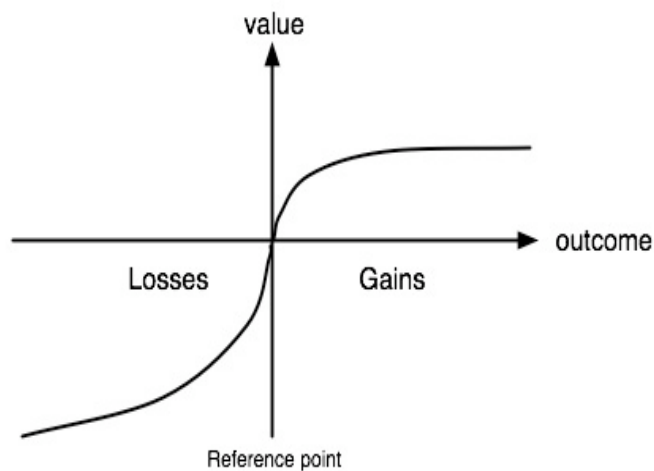
¹ Voir Baron (1988), p 268 ; Levy (1992) et Peterson (2009), p 290.

² Cela pourrait aussi expliquer des phénomènes comme celui-ci : un enfant perd ou brise un de ses jouets accidentellement et on le lui remplace immédiatement, mais ce remplacement ne parvient pas à satisfaire l'enfant autant que la perte l'a affecté. Mais dans ce cas-là intervient ici l'effet de dotation (« endowment effect ») dont nous dirons quelques mots plus loin.

³ Cela explique aussi les phénomènes de surenchérissment dans les pertes et de déchéances des gros joueurs. À ce sujet on peut se reporter au classique de Dostoïevski : *Le joueur*.

- 2) Elle est concave pour les gains et convexe pour les pertes. Rendant ainsi compte de l'aversion au risque dans le domaine des gains et de la recherche des risques dans le domaine des pertes.
- 3) Sa pente est plus accusée pour les pertes que pour les gains, expliquant par là le fait que les pertes pèsent plus lourd que les gains.

Ces trois caractéristiques sont responsables de la forme de la fonction :



+ Le coefficient de pondération.

Comme nous l'avons vu dans les parties précédentes, l'idée d'introduire un coefficient de pondération sur les probabilités était déjà présente dans les articles d'Allais¹ et Edwards², et, comme le remarque ce dernier, la première tentative de formalisation de ce phénomène doit être ramenée à Preston/Baratta (1948)³. Cependant aucune des théories que nous avons étudiées n'était arrivée aussi loin dans la compréhension de ce phénomène. Dans la TP, le coefficient de pondération va être doté d'une capacité explicative jamais atteinte jusqu'à présent. D'autre part, ce coefficient est le deuxième élément – après la fonction de valeur –

¹ Ce dernier était arrivé très près de la formulation de la TP dans son article de 1952 et avait eu l'intuition très nette et précise du phénomène de pondération.

² Notamment Edwards (1954), p 393.

³ Certains auteurs comme Wu/Zhang/Gonzalez (2004), font aussi remonter cette idée à Mosteller/Nogee (1951).

que nous allons utiliser pour reconstruire le concept de volonté. Nous allons, dans notre troisième chapitre, tenter de généraliser les effets de la pondération au temps et à d'autres éléments de la prise de décision comme l'opération de catégorisation qui se déroule pendant la phase d'édition¹.

La TP fait l'hypothèse que chaque probabilité est accompagnée d'un coefficient de décision (« decision weight »²) qui est multiplié par la valeur de la perspective. Comme le remarquent K/T : « Decision-weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences in the Ramsey-Savage approach »³. Le coefficient en question est donc obtenu par la méthode du pari telle qu'appliquée par les deux auteurs, la différence étant que K/T utilisent le concept de perspective et que les probabilités sont objectives. D'où cette question : dans quelle mesure ce coefficient ne peut-il alors être considéré comme une fonction qui rendrait la probabilité subjective? Tout simplement parce que le coefficient en question ne respecte pas les axiomes du calcul des probabilités⁴.

Dans une optique classique, à savoir celle de la TUA, on peut évaluer les préférences d'un sujet en lui proposant un ticket de loterie ou un pari dont les gains sont pondérés par des probabilités, subjectives ou objectives. Mais qu'en est-il si par exemple pour une loterie dépendant du lancer d'une pièce, la pondération de l'événement pile ou face (.50) est légèrement inférieure à la probabilité réelle ? C'est exactement ainsi que fonctionne le coefficient de pondération, il mesure : « the impact of events on the desirability of prospects, and not merely the perceived likelihood of these events »⁵.

Les auteurs vont ensuite passer à l'analyse des caractéristiques de π qui relie les coefficients de décision aux probabilités. Tout d'abord π est une fonction croissante de p et sa valeur est équivalente à p pour les valeurs extrêmes : 0 et 1⁶. Ce qui signifie que comme en TUA les événements nuls ne sont pas pris en compte. La première caractéristique du coefficient

¹ La catégorisation est l'opération par laquelle le SRD pondère la situation.

² K/T utilisent le concept de coefficient de pondération ou de coefficient de décision. Les deux sens ne sont pas absolument équivalents. Le coefficient de décision désigne la pondération d'une option particulière, alors que le coefficient de pondération désigne le phénomène général de pondération des probabilités.

³ K/T (1979), p 34.

⁴ Respect qui est la condition de possibilité de la rationalité elle-même chez Ramsey et Savage.

⁵ K/T (1979), p 35.

⁶ Pour une bonne analyse de ce coefficient, voir Baron (1988), p 263-267 ; K/T (1992) ; et Tversky/Wakker (1995). Le deuxième chapitre de *Choices, Values and Frames* est consacré à π .

de pondération est sa sous-additivité (« subadditivity »)¹ par rapport à p pour les petites probabilités, c'est-à-dire que : $(rp) > r\pi(p)$ pour $0 < r < 1$ ². Si nous revenons aux préférences exprimées par les sujets dans les problèmes **8** et **8'**, la sous additivité conduit aux inégalités suivantes :

$$\frac{\pi(.001)}{\pi(.002)} > \frac{v(3000)}{v(6000)} > \frac{1}{2}$$

Ce qui signifie que le rapport des termes (0,5) bien qu'étant identique partout est perçu comme inégal du fait de la pondération de la faible probabilité par π et des gains possibles par v ³. Il en va de même pour le problème **8'** avec des perspectives négatives. Il faut noter que cette caractéristique ne concerne que les faibles probabilités.

Toujours dans le domaine des faibles probabilités : les auteurs émettent l'hypothèse que les petites probabilités sont généralement surévaluées. Soit le problème suivant :

P14 : Choisir entre (5000, .001)	(5)
N = 95 [72]*	[28]
P14' : Choisir entre (-5000, .001)	(-5)
N = 72 [17]	[83]*

Dans ce cas de figure, les sujets émettent une préférence pour la loterie dans le domaine des gains et une préférence pour la perte sèche dans le domaine des pertes⁴. Ce qui confirme les résultats acquis plus haut : pour les petites probabilités recherche du risque dans le domaine des gains, et aversion au risque dans le domaine des pertes. Cela nous donne un élément de réponse et d'explication pour le problème récurrent de l'achat simultané par les sujets d'assurances et de billets de loterie. Comme le remarquent K/T, dans le cadre des pertes,

¹ Cette question est traitée à fond par Tversky/Wakker (1995), p 1260 et suivantes. Nous renvoyons aussi à leur graphique du coefficient de pondération qui a le mérite de bien mettre en avant la caractéristique de sous additivité ; voir aussi Prelec (2002), p 72-73.

² Dit autrement, pour deux probabilités p et q telles que $f(p+q) \neq f(p) + f(q)$ on a $f(p+q) > f(p) + f(q)$, ce qui peut tout simplement mener à deux évaluations différentes pour une même perspective présentée de deux façons.

³ Nous reviendrons plus tard et plus longuement sur les effets conjugués des deux coefficients.

⁴ Une décomposition du problème permet de voir comment le coefficient de pondération fonctionne. Pour le problème **14** on a : $\pi(.001)v(5000) > v(5)$ donc, en faisant passer le gain de l'autre côté de l'inégalité, cela nous donne : $\pi(.001) > v(5) / v(5000)$, or le rapport $v(5) / v(5000) = 0,001$ ce qui signifie en fin de compte que $\pi(.001) > 0,001$.

les sujets préfèrent : « a small loss, which can be viewed as the payment of an insurance premium, over a small probability of a large loss »¹.

Une autre caractéristique très importante, puisqu'elle permet d'expliquer les incohérences relevées dans les problèmes 1 et 2 ainsi que d'expliquer le paradoxe d'Allais, est la sous-certitude (« subcertainty »), elle concerne cette fois l'addition des probabilités entre elles. On sait que dans l'axiomatique classique des probabilités, l'addition de deux événements, ou d'un nombre quelconque d'évènements doit être égale à un. Or il apparaît que le coefficient de pondération a la propriété suivante : pour tout $0 < p < 1$ nous avons : $\pi(p) + \pi(1-p) < 1$ ²

Si nous prenons par exemple le problème 1 et le problème 2 nous avons les inégalités suivantes, pour **P1** :

$$v(2400) > \pi(.66)v(2400) + \pi(.33)v(2500)$$

C'est-à-dire, en soustrayant $[1 - \pi(.66)]$ partout,

$$[1 - \pi(.66)]v(2400) > \pi(.33)v(2500) \text{ soit : } \pi(.34)v(2400) > \pi(.33)v(2500)$$

Et pour **P2** :

$$\pi(.33)v(2500) > \pi(.34)v(2400)$$

Si nous combinons les préférences exprimées dans les deux problèmes nous obtenons l'expression suivante :

$$[1 - \pi(.66)] > \pi(.34) \text{ or, du fait de la sous-certitude : } \pi(.66) + \pi(.34) < 1.$$

Nous avons ici à faire à deux effets de π , à gauche il crée une inégalité entre deux termes identiques mais qui sont pondérés de façon différente, et à droite la somme des deux options complémentaires n'est plus additive à un, justement du fait de la pondération, il y a

¹ K/T (1979), p 36.

² C'est précisément cette caractéristique qui fait que le coefficient de pondération ne peut être considéré comme une mesure probabiliste.

donc de ce fait une sous-estimation, et c'est aussi ce qui se passait dans le paradoxe d'Allais : l'option sûre est surestimée par rapport aux options complémentaires. Si nous prenons les préférences exprimées dans ce cas là nous trouvons que le même mécanisme est à l'œuvre et que : $\pi(.89) + \pi(.11) < 1$ ¹. Ce qui permet à K/T de décrire une composante essentielle concernant la décision des sujets en contexte d'incertitude : « The sum of weights associated with complementary events is typically less than the weight associated with the certain event »².

La sous-certitude mène donc elle aussi à des incohérences, car par exemple si nous prenons deux perspectives du type (100, .1) et (100, .50 ; 100, 50), cette caractéristique nous indique que la perspective sûre devrait être préférée, bien que du point de vue des probabilités et des gains les deux options soient parfaitement équivalentes. Cependant, si l'opération d'annulation de la phase d'édition fonctionne, la perspective conçue comme « risquée » sera éliminée.

La dernière caractéristique du coefficient de pondération est la sous-proportionnalité (« subproportionality »), elle permet d'expliquer l'effet du rapport commun. Il s'agit ici de formaliser, au moyen de l'intervention de π , le fait que les préférences peuvent être altérées par l'introduction d'un facteur probabiliste identique dans les options. Par exemple si la perspective (x, p) est équivalente à la perspective (y, pq) , l'ajout d'un facteur probabiliste r tel que $0 < p, q, r < 1$ ne devrait pas changer l'ordre des préférences et (x, pr) devrait être équivalent à (y, pqr) . Par ailleurs, le rapport entre les deux options reste identique :

$$\frac{(x, p)}{(y, pq)} = \frac{(x, pr)}{(y, pqr)}$$

Or du fait de l'action du coefficient de pondération nous obtenons, pour les probabilités :

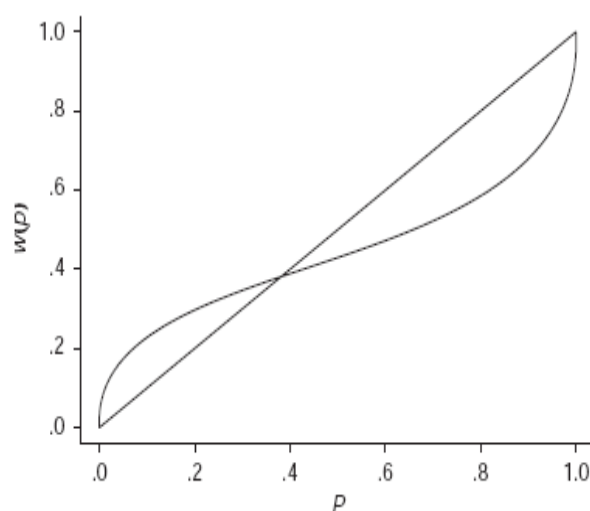
$$\frac{\pi(p)}{\pi(pq)} \leq \frac{\pi(pr)}{\pi(pqr)}$$

¹ Par contre comme nous l'avons vu plus haut : $\pi(1) = 1$.

² K/T (1979), p 36. Ce qui signifie aussi que la somme des parties n'est pas équivalente à la totalité, il y a un vide quelque part.

Ce qui nous indique que pour un rapport fixe de probabilités, le coefficient de pondération correspondant se rapproche plus de l'unité quand les probabilités sont faibles que quand elles sont élevées¹. L'égalité du rapport n'est pas perçue par les sujets comme l'indique la deuxième version du paradoxe d'Allais ainsi que les problèmes **3 à 8** de l'article.

Les effets croisés de la sous-proportionnalité et de la surestimation des petite probabilités expliquent la forme du coefficient de pondération (voir schéma ci-dessous), c'est-à-dire la sous-additivité pour les petites probabilités, qui est peut-être la caractéristique la plus importante de π . Ces caractéristiques produisent un coefficient dont le graphe est le suivant :



Graphique qui permet de voir comment le coefficient se comporte par exemple pour les petites probabilités, et comment les sujets sous-évaluent les hautes probabilités quand elles ne sont pas extrêmement élevées puisque dans ce cas-là elles sont assimilées à la certitude, les probabilités extrêmement faibles sont elles assimilées à l'impossibilité. La forme du coefficient de pondération permet aussi d'expliquer pourquoi un changement d'une probabilité nulle à la possibilité, et d'une possibilité à la certitude est toujours surévalué par rapport à un changement qui se situerait dans le milieu de l'intervalle². Pour illustrer ce phénomène les auteurs empruntent à Richard Zeckhauser l'exemple de la roulette russe³ dont la question est la

¹ Voir Prelec (2002), p 71, en particulier sa formalisation de l'effet du rapport commun à partir de l'exemple original d'Allais.

² Voir Baron (1988), p263-265 ; Wakker/Tversky (1995).

³ Binmore (2009), p 44-46, traite cet exemple dans toute son extension et explique notamment pourquoi, en faisant appel au calcul, la première alternative est plus rationnelle que la seconde.

suivante : contraints de jouer à la roulette seriez-vous prêts à payer plus cher pour enlever une balle qui ferait passer le nombre de 4/6 à 3/6 ou pour retirer un projectile qui ferait passer le nombre de balles de 1/6 à 0/6. Bien que l'exemple soit un peu artificiel, il est bien évident que tous les sujets se portent spontanément vers la deuxième alternative. Le passage de la possibilité à l'impossibilité ou de la possibilité à la certitude ne doit pas juste être pensé comme un changement sur une échelle continue bornée par 0 et 1, il s'agit en fait de véritables changements de catégorie dont l'impact sur l'existence des sujets peut être dramatique.

Les auteurs passent ensuite au problème de la dominance stochastique que la TP ne prend pas en compte et qui pourrait donc constituer une objection contre la présente théorie. Seulement, bien que problème soit important¹, ce serait oublier l'opération de détection de la dominance dans la phase d'édition, qui a pour effet d'éliminer d'entrée de jeu les perspectives dominées au profit de celles qui les dominent.

c – Les effets croisés des deux coefficients et les extensions de la TP.

Comme nous l'avons vu, la sous-certitude permet d'expliquer les problèmes 1 et 2 ainsi que le paradoxe d'Allais, ce qui n'est pas négligeable. Mais ce type de violation n'advient que dans les cas où : « The v -ratio of the two non-zero outcomes is bounded by the corresponding π -ratio »². *C'est donc l'action combinée des deux coefficients qui produit les violations de la TUA*, que ce soit pour l'effet de la conséquence commune, ou pour l'effet du rapport commun. *L'action des deux coefficients produit une pondération globale de la perspective*. Ce qui est à la source des violations de la TUA ce sont donc des éléments issus de la façon dont les sujets se *représentent* les gains et les probabilités.

Comme le montrent les deux inégalités de la page 39, dans les problèmes 1 et 2 et les problèmes 7 et 8, le rapport des gains des deux issues pondérées par la fonction de valeur se trouve limité par les rapports des probabilités qui sont pondérées différemment alors que leurs valeurs sont égales. *La pondération produit des inégalités là où les choses sont en fait égales*. C'est cela même qui produit les violations caractéristiques de l'axiome d'indépendance-substitution. Comme le remarquent K/T : « The same analysis applies to other violations of the

¹ Puisque certains auteurs, notamment Quiggin (1982), construiront leurs théories, dites alors « dépendante du rang » (« rank dependant »), pour pallier à cette insuffisance de la TP. D'ailleurs K/T dans la TCP de 1992 feront en sorte d'intégrer le phénomène de la dominance stochastique

² K/T (1979), p 39.

substitution axiom, both in the positive and in the negative domain »¹.

D'une façon plus générale, l'interaction des deux coefficients produit un ensemble de phénomènes caractéristiques qui peuvent nous aider à mieux comprendre le tableau des quatre modèles de comportements par rapport au risque. Par exemple le travail commun de la sous-évaluation des probabilités moyennes ou hautes et de la concavité de la fonction de valeur² a pour conséquence d'augmenter de façon sensible l'aversion pour le risque. Dans le domaine des pertes, l'action de π minore la pondération des perspectives risquée, créant ainsi une recherche du risque. Dans le domaine des probabilités moyennes et hautes, nous pouvons dire que l'action des deux coefficients se renforce.

Si nous passons maintenant au-dessous du point de transition entre les probabilités moyennes et les faibles probabilités nous observons que la surestimation des probabilités augmente l'estimation de la valeur des perspectives positives et l'estimation de la valeur négative des perspectives négatives. Ce qui a pour conséquence une majoration de la recherche du risque dans le domaine des gains et une aversion aux risques dans le domaine des pertes. Or cela est contraire aux effets de la fonction de valeur. Dans ce cas-là, les deux coefficients ne se renforcent pas l'un l'autre. Le fonctionnement conjoint des deux coefficients permet ainsi d'expliquer du même coup l'achat simultané d'assurance et de billets de loterie par le même sujet : les deux phénomènes sont liés à l'action de π et de ν dans le domaine des probabilités faibles.

En ce qui concerne le rôle de la fonction de valeur la TP prédit plusieurs éléments importants, quelques remarques s'imposent donc. Tout d'abord le point de référence peut être positionné par rapport aux gains et aux pertes du sujet non plus en les rapportant au statu quo mais à un projet ou un objectif à atteindre³. Auquel cas ce sera en fonction de ce point de référence que les gains et les pertes seront codés pendant la phase d'édition. Ce phénomène nous laisse entrevoir la possibilité d'une fonction de valeur dynamique, et par là même d'un système de référence décisionnel dynamique. Cependant dans le contexte de la TP, la dynamique en question peut avoir des conséquences assez graves.

Le fait que le point de référence soit susceptible de se déplacer par rapport à l'avoir

¹ K/T (1979), p 39.

² Qui a pour effet de favoriser les perspectives sûres par rapports aux perspectives risquées.

³ Voir K/T (1979), p 40. Ce processus est à rapprocher de l'heuristique d'ancrage et d'ajustement dans sa façon de fonctionner, cependant on ne peut affirmer que ce soit un biais.

actuel du sujet peut entraîner un changement dans l'ordonnancement des préférences, et cela peut avoir des conséquences en particulier au niveau des pertes, par exemple si le sujet n'a pas adapté son système de référence à des pertes récentes : « The negative translation of a choice problem, (...), increases risk seeking in some situations »¹. Ce qui indique qu'un sujet qui n'est pas en adéquation avec l'état réel de ses pertes² sera porté à accepter des paris qu'il aurait refusés en temps normal. C'est le cas par exemple chez les joueurs pathologiques qui acceptent n'importe quel pari sous prétexte de se « refaire », ce faisant les paris en question ne sont plus du tout codés comme des perspectives risquées qui pourraient aggraver la situation, mais comme de pures possibilités, d'où l'aspect négatif est complètement éliminé. Le gain possible, même avec une probabilité infime est complètement surévalué. C'est encore une fois la façon dont les sujets se représentent la situation qui est centrale, car la phase d'édition est complètement déterminée par la façon dont le sujet se représente la réalité décisionnelle, d'autre part, la fonction de valeur est tributaire de ce genre d'influence de la capacité représentative. Il s'agit donc ici d'un phénomène de décalage entre l'avoir actuel du sujet et sa fonction de valeur qui peut augmenter la recherche du risque. Comme le remarquent K/T pour terminer : « The location of the reference point, and the manner in which choice problems are coded and edited emerge as critical factors in the analysis of decisions »³.

K/T concluent leur article en proposant quelques axes de recherches en ce qui concerne le devenir de la TP. La première remarque concerne la généralisation de la théorie à un nombre quelconque d'issues, dont les auteurs soutiennent ici qu'elle va de soi. Or la lente construction de la Théorie Cumulative de la Perspective (1992)⁴ montrera au contraire que la généralisation n'allait pas du tout de soi, puisque la formalisation d'un nombre quelconque de perspectives, le passage contexte de risque à l'ignorance et l'intégration de la dominance stochastique allaient nécessiter un renouvellement considérable des mathématiques de la TP, par exemple l'utilisation de la théorie des capacités de Choquet à la place des probabilités.

¹ K/T (1979), p 41.

² Cette inadéquation peut provenir de deux facteurs, qui peuvent agir seuls ou combinés : la fonction de valeur est décalée par rapport à l'avoir du sujet, les pertes ne sont pas codées comme telles mais de façon biaisée dans la phase d'édition.

³ K/T (1979), p 42.

⁴ Voir K/T (1992), et Wakker (2010), le livre de Wakker constitue la meilleure « introduction » à la TCP. Sur le concept de Capacité voir Choquet (1955). Des théories cumulatives avaient justement été proposées, par Choquet, Schmeidler (1989) et surtout Quiggin (1982). Pour une revue générale des théories dépendantes du rang, voir R. Sugden, in Barberà /Hammond/Seidl (Eds.) (2004), p 712-723.

À notre sens et pour la problématique générale de notre travail, l'apport principal de la TCP est le dédoublement du coefficient de pondération. Partant du principe établi dans la TP selon lequel les pertes pèsent plus lourd que les gains, K/T vont avoir l'idée que le coefficient de pondération pour les perspectives positives n'aura pas exactement la même forme que pour les perspectives négatives¹.

Deuxième remarque, les auteurs font l'hypothèse que la TP pourrait être généralisée à des issues non monétaires, telles que la qualité de la vie, et les décisions politiques², en particulier en ce qui concerne l'action croisée du coefficient de pondération et la fonction de valeur. Nous proposerons aussi sous peu une généralisation philosophique de la TP, généralisation qui nous permettra de repenser et de redéfinir le concept de volonté.

Enfin, K/T envisagent que le coefficient de pondération puisse être appliqué directement aux événements eux-mêmes lorsque les probabilités ne sont pas claires ou ne sont pas données explicitement. Il eut été intéressant de confronter cette idée avec le paradoxe d'Ellsberg par exemple, afin de voir si la pondération des événements eux-mêmes pouvait donner une explication du phénomène. Bien que l'essai n'ait pas été fait, l'idée d'une pondération des événements laisse penser que K/T avaient entrevu la possibilité d'appliquer un coefficient de pondération à d'autres éléments du contexte décisionnel que les probabilités. La question qui émerge alors est la suivante : pour qu'un coefficient de pondération soit applicable à d'autres éléments que les probabilités est-il nécessaire que l'élément en question soit isomorphe à ces dernières ? Ce qui est le cas par exemple pour des événements complémentaires. Nous montrerons que le concept de pondération doit être parfois modifié pour être adéquatement généralisé.

D'un point de vue plus général et conclusif, quelques remarques s'imposent avant de passer au troisième chapitre de ce travail. Les critiques de la TUA que nous avons vu se développer d'Allais à L/S ont en commun le fait que dans une situation décisionnelle ou dans un contexte de choix donné, il y a une influence de la capacité représentative³ des sujets sur les

¹ Voir K/T (1992), p 60.

² Kahneman lui-même s'occupera de la première option, Isaac Levy de la seconde.

³ Nous pourrions aussi dire « cognition » au sens large.

probabilités et les issues des alternatives. Le problème de la représentation est récurrent depuis le début de notre chapitre, et nous avons vu à la fin de notre parcours, notamment avec L/S, que plusieurs facteurs d'ordre représentatifs peuvent intervenir et avoir une influence combinée dans la construction des préférences et ce en fonction de la tâche demandée (évaluer monétairement/jouer), et des stimuli en présence (Pari P, ou pari \$). Cette idée était déjà présente dans Allais, qui l'avait partiellement développée et formalisée¹, puis dans Edwards qui avait eu l'intuition très nette de ce que devrait être un coefficient de pondération sur les probabilités afin d'expliquer les incohérences des sujets². Les recherches d'Ellsberg elles aussi avaient montré que c'est la façon dont les sujets se représentent les alternatives³ qui met en faillite la distinction classique entre les contextes inaugurée par Knight et le PCS de Savage.

Or aucune des théories que nous avons étudiées dans notre premier chapitre ne mentionne ou ne problématise quoique ce soit qui ait un rapport avec la façon dont le sujet se représente les options, ce sont de pures structures mathématiques dont les prétentions normatives sont très élevées. Cela ne signifie pas que les théories en questions sont inutiles⁴ ou inutilisables, mais que leur coût d'utilisation est trop important lorsqu'on le ramène aux capacités restreintes des sujets, d'autant plus – et c'est ce que démontrent les contestations empiriques – que leur utilisation est très loin d'être spontanée⁵. Quand nous confrontons la TUA à ses réfutations empiriques, ce qui émerge c'est l'absence d'élaboration du rapport des sujets aux options, aux issues et aux probabilités. Lorsque nous parlons ici de « rapport » nous parlons en fait de la façon dont les sujets intériorisent et codent les décisions. Or ce rapport des décideurs à leurs décisions s'exprime par le phénomène général de la pondération et de la distorsion de l'information : les sujets ne se représentent pas les perspectives et les décisions telles qu'elles sont en fait. La mise en présence des axiomes de la TUA avec les critiques que nous venons d'étudier permet de faire apparaître deux éléments centraux que la TP récapitule

¹ Allais était arrivé très proche de la TP. Voir en particulier Allais (1952), p 48.

² Comme nous l'avons remarqué, cet auteur avait même eu l'intuition du dédoublement des coefficients de pondération que formaliser la TCP en 1992.

³ En fait, ici il s'agit plus précisément de la façon dont les sujets perçoivent et se représentent l'information probabiliste, qui dans la théorie classique de la décision détermine à quel contexte nous avons à faire.

⁴ Encore une fois, nous renvoyons à la remarque de de Finetti : le non-respect des axiomes des probabilités, dans le cadre d'un dutch book, peut entraîner des pertes qui sont loin d'être virtuelles...

⁵ Encore une fois, il est utile de remarquer que même si la critique a été poussée, notamment par Allais (1953), jusqu'à réfuter l'axiome de réduction des loteries composées parce qu'il efface un élément aussi important que le plaisir potentiel pris au jeu, il y a un axiome qui résiste particulièrement bien : l'axiome d'Archimède, qui malgré les critiques conserve toute sa portée descriptive et normative.

et formalise : le coefficient de pondération et la fonction de valeur, qui sont tous deux à verser au registre de la représentation subjective, des probabilités pour le premier, et du codage des gains et des pertes pour le second. Cependant, la TP reste encore pour nous trop conservatrice et ancrée dans la lignée de la TUA¹. Son mérite majeur est d'avoir montré tout d'abord comment le sujet intériorise les perspectives en les éditant, et ensuite comment la perspective éditée est évaluée par les deux coefficients.

La TP a eu aussi le mérite de mettre au jour un autre phénomène très important, qui est lié à l'action combinée de la phase d'édition et de la fonction de valeur : l'effet de cadrage (« framing effect »²). Effet de cadrage qui peut être entendu de deux façons : tout d'abord comme la construction de la décision par le sujet, du fait du jeu combiné de la phase d'édition et la phase d'évaluation, et d'autre part comme la façon dont le même sujet peut être affecté, dans son choix, par la façon dont on lui présente les alternatives. Par exemple, deux perspectives identiques peuvent être éditées et évaluées de façons différentes selon qu'on l'on présente l'une comme une loterie et l'autre comme un jeu à deux étages. Ce qui signifie que la capacité décisionnelle des sujets peut être manipulée, notamment en contexte économique.

En ce qui concerne la suite de notre travail, l'analyse de la TP nous a permis de faire émerger, à partir des critiques de la TUA les deux éléments centraux sur lesquels nous allons élaborer notre propre théorie de la volonté, puisque nous émettons ici l'hypothèse que le travail de la volonté consiste en un ensemble de processus que nous allons explorer comme les opérations du Système de Référence Décisionnel (SRD), dont les trois fonctions principales sont l'établissement de la visée, l'édition et l'évaluation. Les concepts de coefficient de pondération et de fonction de valeur seront donc généralisés et adaptés à une théorie d'ordre plus philosophique. Il sera par exemple question de réfléchir sur la façon dont les sujets pondèrent le temps, en analysant des concepts comme ceux de délai ou d'urgence notamment.

Par ailleurs le travail de la volonté ne se situe pas pour nous dans un contexte, mais dans un environnement décisionnel qu'il faudra définir, justement en déconstruisant le concept classique de contexte décisionnel, trop étroit puisqu'il fait seulement appel à une information de type probabiliste. Le concept d'environnement devra donc être défini d'une part en relation avec les capacités restreintes des sujets, d'autre part en relation avec les concepts d'horizon

¹ Même avec la TCP de 1992. Il conviendra donc pour nous de faire sortir les deux coefficients de ce modèle computationnel.

² Voir K/T (1986).

(spatial et temporel¹) et de monde. Il nous faudra donc construire les outils conceptuels qui nous permettront de relier entre eux les différents éléments théoriques que nous avons mis au jour, et à partir desquels nous entendons élaborer nos propres hypothèses.

¹ Mais aussi existentiel, la mort constituant ici l'horizon absolu des événements, mais non des conséquences. C'est une distinction très importante.

Chapitre III

Le Système de Référence Décisionnel :

La décision dans l'élément de la représentation.

Introduction.

Le chapitre que nous commençons ici est une analyse partielle¹ du fonctionnement du SRD², ou Système de Référence Décisionnel. Ce système s'articule autour de trois opérations dont deux sont directement issues de la Théorie de la Perspective de Kahneman et Tversky : *établissement de la visée* ou de *l'environnement décisionnel*, *édition*, *évaluation*. Le SRD est le concept philosophique par lequel nous entendons reconnecter le décidable avec le volontaire, qui avait été banni du champ d'investigation de la TUA et de sa critique par la psychologie. Comme nous l'avons noté à plusieurs reprises nous définissons la volonté comme *la décision dans l'élément de la représentation*. Le SRD est donc équivalent à *l'acte* de la volonté, il est la volonté s'actualisant de façon processuelle dans une décision particulière.

Le SRD est tributaire de façon absolue d'un système complémentaire, que nous baptisons le Système de Référence Général (SRG), ce système contient en lui tout les éléments représentationnels et cognitifs qui assurent au sujet son irréductible singularité, à savoir sa capacité d'être affecté, ses connaissances, ses croyances, ses valeurs et ses souvenirs, tout ce qu'il a acquis au cours de son histoire, de son expérience et qui le définit comme *ce sujet-ci*. La singularité du sujet et son individualité sont déterminées par les relations que les éléments du SRG entretiennent entre eux. Le SRD est lui aussi inclus dans le SRG, dont il est un sous-système, ce qui signifie que tous les autres éléments du SRG peuvent avoir des influences diverses sur son travail et son fonctionnement, notamment les affects et les émotions, puisque comme nous allons le voir, la première évaluation que le SRD opère est une évaluation affective³ de la situation.

Il s'agit donc ici, comme nous le remarquons avec plus de détails dans notre introduction générale d'une théorie qui repose sur l'exploitation et la reconstruction

¹ Notamment du fait que les analyses de la phase d'évaluation ne seront pas étudiées intégralement. En fait pour être parfaitement exhaustif il faudrait élaborer philosophiquement tous les items qui sont inclus dans le SRG, car tous sont susceptibles d'être déployés comme c'est le cas pour le SRD. Une série d'études de détails, à venir, pourraient venir combler cette lacune. Lacune redoublée par le fait qu'il faudrait détailler le fonctionnement du SRD pour toutes les catégories de décisions.

² Pour plus de clarté, se reporter au schéma qui clôt cette introduction, p 219,

³ Il conviendra donc de penser le lien du SRD aux affects, et de calibrer le rôle réel que les émotions peuvent avoir dans la prise de décision, en se libérant du préjugé classique selon lequel elles ne sauraient être qu'un obstacle au vouloir. Nous situerons donc nos réflexions en continuité avec celles de Zajonc (1980), de Sousa (1990), et de Damasio (1995), mais aussi celles plus récentes de Tappolet (1995) et (2003).

philosophique des principaux acquis de la TP de 1979. Revenons sur ce qui légitime cette opération. La TP a mis au jour cinq faits fondamentaux de la prise de décision en contexte de risque qu'il faut ici rappeler : la phase d'évaluation, la phase d'édition, et à l'intérieur de cette phase la TP décrit comment la fonction de valeur et le coefficient de pondération s'appliquent aux perspectives, le premier aux valeurs des issues à partir de l'avoir présent du sujet, le deuxième aux probabilités à partir d'une pondération subjective liée à la façon dont le sujet se *représente* les probabilités. Enfin, phénomène très important pour nous, les sujets sont d'une façon générale aversifs aux pertes. Or nous nous sommes aperçus au cours de nos réflexions que tous ces éléments, repensés de façon philosophique et réorganisés pouvaient trouver leur place dans une structure conceptuelle plus vaste, départie des limitations originales de la TP. Mais pour ce faire, il fallait passer d'une théorie à caractère *formel* à une théorie à caractère plus conceptuel, et ce faisant plus générale. Il fallait prolonger philosophiquement la TP et ajouter aux deux phases qu'elle distingue, une phase préalable d'établissement de la visée.

Mais, d'une façon plus précise encore, l'idée même d'un système de référence décisionnel est directement issue du concept de la fonction de valeur, dont il est la généralisation *naturelle* du point de vue de la philosophie, car il représente – du fait du lien essentiel qui l'unit au SRG – l'*état présent*¹ par rapport auquel toute situation décisionnelle est construite et évaluée. Le SRD représente *le point de vue du sujet* dans la sphère décisionnelle, et c'est aussi pour cela que, par métaphore, nous parlons de « système de référence »², et surtout ce qui compte ici, c'est que ce point de vue soit absolument singulier.

Un autre fait central que met en avant la TP, c'est la pondération des probabilités par un coefficient de pondération : $\pi(p)$. Coefficient dont le but est de formaliser la façon dont les sujets se représentent les probabilités objectives, et par là même les distordent. Or, comme nous le savons, la plupart des alternatives que nous utilisons et produisons usuellement ne sont pas accompagnées de mesures probabilistes objectives, mais d'évaluations qui ne respectent pas les axiomes du calcul des probabilités, et que capture le concept de « chance de » qui va être élaboré dans ce chapitre. Cependant il appert que même si la situation ne comporte en elle aucun élément probabiliste une pondération se produit quand même par le biais de l'opération

¹ Ce qui signifie aussi que nous allons devoir construire le concept d'état présent à partir de celui d'avoir présent que promeut la TP.

² Métaphore qui trouve son origine dans la théorie de la relativité d'Einstein, où les systèmes de référence peuvent être conçus comme autant de points de vue sur la réalité physique.

de *catégorisation* qui consiste pour le SRD à évaluer la situation selon qu'elle est habituelle non-critique, habituelle critique, inhabituelle non-critique, inhabituelle critique ou exceptionnelle. Cette opération sert à penser comment le sujet pondère la situation, même lorsque les probabilités sont absentes.

Enfin, un dernier élément très important est à prendre en compte, il s'agit de l'aversion au risque dont l'une des conséquences est une *préférence marquée des sujets pour le statu quo*. Cette surévaluation de l'état présent nous a semblé pouvoir expliquer un certain type d'akrasie, liée au fait que le SRD peut aussi être une source d'inertie et de résistance, notamment parce que ses opérations usuelles oscillent autour de certaines habitudes, le SRD a une tendance conservatrice assez marquée, bien qu'il puisse aussi intégrer et produire de nouvelles stratégies. Comme l'existence se stabilise la plupart du temps autour de certains pôles habituels, le SRD va suivre cette tendance et axer son fonctionnement – qui se doit de toujours être le plus économique possible – sur les mêmes stratégies décisionnelles. Mais il faut aussi que le SRD ait en lui la capacité d'évoluer, ne serait-ce que pour être en adéquation avec l'évolution générale du SRG et de l'existence subjective.

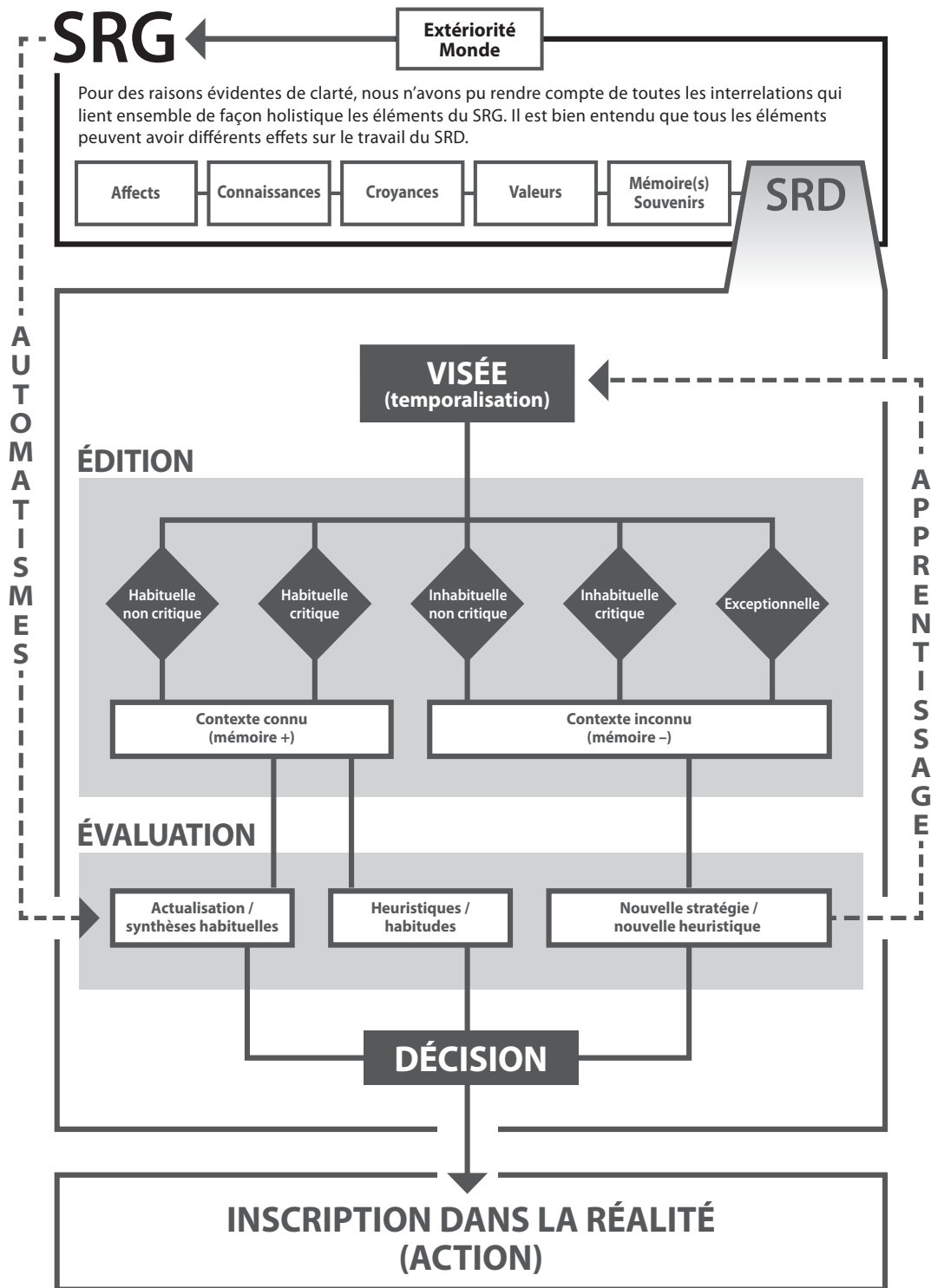
Dans cette reconstruction nous tentons de rester le plus proche possible de l'idée de la TP, mais aussi de rester le plus proche possible de l'expérience et de la façon dont se déroulent *concrètement* les prises de décision. Ce fonctionnement est en fait assez simple, reprenons-le d'une façon synthétique en suivant le schéma de la page 219. Le SRG est connecté au monde extérieur, il est le système de référence général et contient l'état affectif du sujet, ses croyances, ses valeurs, ses connaissances et ses souvenirs, ainsi que le SRD. Il est important que le SRG maintienne une connexion constante entre l'intériorité et le flux informatif venant du monde extérieur, pour deux raisons : dans certains contextes à définir les alternatives sont présentées au sujet de l'extérieur, par exemple le choix d'un menu dans un restaurant, et ne sont pas construites par le SRD. C'est aussi de cette façon que les alternatives sont présentées au sujet dans la Prospect Theory. La décision s'opère donc dans le SRD mais à partir d'un matériau qui doit être assimilé et reconstruit (framing effect). Deuxième chose : l'information nécessaire à la construction des alternatives peut venir de l'extérieur, et c'est par le SRG qu'elle va transiter avant de devenir disponible pour le SRD.

Le fonctionnement général du SRD est le suivant : la première opération consiste à établir la visée, que nous allons construire philosophiquement comme l'environnement

décisionnel, c'est-à-dire le lieu métaphorique dans lequel la décision et le but vont pouvoir s'inscrire. À cette opération qui est aussi temporalisation succède l'édition des alternatives – au moyen d'*heuristiques d'éditions* – en deux phases : la catégorisation et la contextualisation qui est reliée au rôle que peut jouer la mémoire. Ces deux phases sont préalablement conditionnées par une édition affective et une édition doxique. Vient ensuite l'évaluation des alternatives, qui dépendra du type de catégorie dans laquelle les alternatives ont été éditées. Dans cette phase le SRD utilise ce que nous analyserons sous le terme d'« *heuristiques d'évaluations* », mais il peut se trouver dans des situations où il doit produire une stratégie nouvelle à partir de son propre fonds, stratégie qui sera par la suite, si elle s'est avérée pertinente, intégrée dans la mémoire. C'est là le processus d'apprentissage du SRD.

Deux éléments restent à souligner. Pouvons-nous dire ici que nous tournons le dos de façon définitive à la TUA et à la théorie classique de la décision ? Non, pour une raison simple, c'est que cette construction théorique sera pour nous l'occasion de nous rendre compte que certains axiomes issus de la TUA gardent leur validité comme l'axiome de continuité et le PCS, ainsi que l'idée de Savage selon laquelle le « monde » qui doit être pris en compte par le sujet pour une décision particulière doit être taillé à la mesure de cette décision¹, c'est-à-dire doit être le plus précisément limité et délimité possible. Deuxième remarque, toutes ces analyses se situent dans le cadre de l'hypothèse d'une *rationalité limitée* élaborée par Simon (1955), et dont Gigerenzer (1996) et Gigerenzer/Brighton (2009) feront plus tard la promotion intensifiée. C'est aussi pour cela que nous allons commencer le chapitre par une analyse des travaux de Gigerenzer et Al., travaux qui vont nous donner en quelque sorte le terrain sur lequel nous allons construire nos propres hypothèses.

¹ Voir Savage, (1954), p 9.



A – Rationalité, contexte et environnement décisionnel.

1) La double limitation de la rationalité.

La TP, bien que représentant un progrès décisif du point de vue du réalisme par rapport à la TUA – grâce à la distinction des phases d'édition et d'évaluation et l'introduction du coefficient de pondération et de la fonction de valeur – partage avec cette dernière un défaut assez grave : aucune des deux théories ne prend en compte le fait que d'une façon assez évidente la rationalité des sujets est *limitée* par différents facteurs¹. Pourtant, dès 1956 Herbert Simon envisageait la possibilité qu'en contexte décisionnel, la rationalité devait être limitée de façon exogène et endogène : d'une part du fait de la structure de l'environnement, d'autre part du fait de la limitation des capacités subjectives². Pour Simon il fallait penser les décisions humaines en termes de satisfaction, de niveau d'aspiration³, et non en termes d'optimisation ou de maximisation, puisque cette dernière présuppose un outillage intellectuel dont la portée réelle est délicate à délimiter et dont l'utilisation par les sujets fait question. Cette double contrainte, rationnelle et environnementale, fera dire à Simon dans un article de 1990, que d'un point de vue général : « Human rational behavior (and the behavior of all physical symbol systems) is shaped by a scissor whose two blades are the structure of task environments and the computational capabilities of the actor »⁴. C'est cet auteur qui introduisit le terme de « bounded rationality » dans son article de 1956,⁵ car dans une décision il faut

¹ On pourrait toujours objecter à cet avis que la TUA étant une théorie strictement mathématique n'a pas besoin d'envisager ce genre de limitation. Ce serait le cas en effet si elle n'avait pas une portée descriptive et normative aussi affirmée. De son côté la TP, bien qu'ayant une base empirique assez robuste, reste dans le paradigme de la TUA et repose donc sur l'idée de maximisation, or c'est justement cette idée qui va être au centre des critiques que nous allons étudier. Les deux théories accordent aux décideurs *trop* de rationalité. Voir à ce sujet Gigerenzer/Todd in Gigerenzer/Todd (Eds.) (1999), p 5-10 ; voir aussi Gigerenzer in Gigerenzer/Selten (Eds.) (2001), p 38-40. Par ailleurs, comme nous le verrons plus tard il convient aussi d'injecter de l'affectif dans le rationnel, car nos décisions dépendent en grande partie de nos affects.

² Voir Simon (1956), p 129-131. Notons que l'introduction du concept d'environnement issu de la biologie (à la différence du concept de contexte issu de l'économie), met en avant de façon directe le problème de la collecte de l'information et de son coût, à ce sujet voir Simon (1955), p 106. Problème qu'avait élaboré et formalisé Ramsey (1926) mais qui sera ignoré dans le modèle N/M et dans la TP. Par ailleurs Savage, dans le cadre de son bayésianisme abordera le problème mais de biais : par le concept de révision des croyances qui nous semble différent de celui de collecte d'information, même s'il est évident que l'information collectée aura une influence sur les croyances. Notons que ce problème était au centre de la critique d'Ellsberg (1961). Les limitations que Simon avait pointées du doigt étaient restées largement ignorées.

³ Ce concept nous rapproche de celui de but ou visée qui viendra bientôt remplacer celui de résultat ou de gains.

⁴ Simon (1990), p 7. Il conviendra ici d'appliquer cette métaphore à la théorie de la décision en reconstruisant le concept d'environnement.

⁵ Concept qui allait refaire surface seulement dans les années 1990 avec les travaux de Gigerenzer (1991) et Gigerenzer/Goldstein (1996). De leur côté Cosmides/Tooby (1996) allaient mettre en relation les limitations de la rationalité humaine avec le caractère adaptatif de cette rationalité. Cette perspective adaptative n'est pas absente des travaux de Simon

tenir compte du rapport du sujet au monde environnant et surtout de la résistance¹ de l'environnement lui-même, mais aussi des outils cognitifs limités dont dispose le SRD pour construire ses décisions. Nous allons étudier la question de la rationalité limitée en nous fondant sur les travaux de Gerd Gigerenzer et son groupe de recherches, travaux qui sont à inscrire dans la droite ligne de ceux de Simon. Nous verrons alors émerger un nouveau concept de rationalité, presque paradoxal, puisque contre toute la tradition occidentale depuis Parménide et Platon, il met au centre du choix rationnel la connaissance que les sujets peuvent produire de façon *spontanée et non déductive*, ainsi que les informations qu'ils peuvent tirer *immédiatement* de leur environnement. Ces analyses vont aussi nous servir à introduire dans notre travail le concept d'environnement décisionnel, dont la délimitation pour une décision donnée est la première tâche du SRD.

Pour Gigerenzer et Al. il va s'agir de montrer que dans le domaine de la décision il existe un usage positif et nécessaire d'un certain type d'heuristiques². Les analyses vont donc s'inscrire en faux par rapport à ce que K/T pensaient de l'utilisation de ces « méthodes » par les décideurs, à savoir qu'elles constituent une défaillance essentielle de la rationalité des sujets qu'il faut tenter d'éviter. Gigerenzer et son équipe vont donc revisiter le concept d'heuristique et le rapprocher de celui d'algorithme³. Pour ce faire il s'agira de comparer au moyen d'expériences les résultats qui sont obtenus en utilisant soit une méthode de computation « optimale » soit une ou plusieurs heuristiques, autrement dit il va être question de faire se mesurer « intuition »⁴ et raisonnement. Comme le remarque Klein (2001), à propos des heuristiques : « I have found it more useful to see them as strengths that permit skillful decision making in field settings »⁵.

Le concept d'heuristique ainsi repensé va donc s'inscrire à la fois contre les vues

(1955), ni de ceux de Gigerenzer, voir par exemple, Goldstein/Gigerenzer (2002), in Gigerenzer (Ed.) (2011), p 79. Elle est par contre absente de la TUA et de la TP dans leurs formulations originales. Cependant Maynard Smith (1982) s'est interrogé sur les relations entre la théorie des jeux et l'évolutionnisme dans le but de formaliser le concept de stabilité adaptative. D'autre part le lien entre TP et théorie de l'évolution a été établi par McDermott/Fowler/Smirnov (2008).

¹ C'est un concept philosophique central pour notre propos : décision et résistance sont analytiquement liées.

² Voir Gigerenzer/Todd in Gigerenzer/Todd (eds) (1999), p 27 ; Todd in Gigerenzer/Selten (eds) (2001), p 52 ; Gigerenzer/Brighton (2009) in Gigerenzer (ed) (2011) p 4. Il s'agit ici d'un renversement de perspective important puisque les décisions fondées sur des procédures heuristiques ne seront plus considérées comme prises en faute par rapport à une théorie normative.

³ Voir, Gigerenzer/Goldstein (1996) in Gigerenzer (ed) (2011) p 37. Puisque les heuristiques en question, par exemple la *recognition*, vont être promues au rang de véritables méthodes de décision.

⁴ Nous utilisons ici le terme d'« intuition » au sens de jugement spontané qui se différencie, voire qui s'oppose au raisonnement et à la computation.

⁵ Klein, in Gigerenzer/Selten (Eds.) (2001), p 114.

pessimistes proposées par K/T dans les années 70, mais aussi contre les théories qui soumettent la rationalité décisionnelle au calcul et à la computation, c'est-à-dire les théories qui reposent sur la maximisation ou sur l'optimisation¹. Puisque l'optimisation (qui inclut la maximisation de l'utilité attendue et la TP) suppose elle aussi une forme de calcul².

Afin de conceptualiser la limitation de la rationalité des sujets en contexte décisionnel, Gigerenzer et al. vont être amenés à proposer une théorie baptisée « Fast and Frugal Heuristics »³ dont nous allons examiner les éléments principaux qui se déploient dans plusieurs directions, notamment une nouvelle définition du concept d'heuristique par la mise au jour de nouvelles « méthodes » heuristiques, et d'autre part une redéfinition de la rationalité qui nous semble plus réaliste que celles que nous avons vues jusque alors⁴ puisqu'elle est basée sur les rapports du sujet à l'environnement qu'il construit⁵.

Dans l'économie générale de notre travail, les analyses de Gigerenzer et Al. vont nous permettre d'approfondir le rôle des heuristiques en contexte décisionnel puisque nous pensons que dans les phases d'édition et d'évaluation le travail du SRD est conditionné par ces méthodes de réponse écologiques et économiques. L'idée centrale de Gigerenzer et Al.⁶ est qu'en contexte décisionnel de simples heuristiques peuvent produire des résultats plus précis que des méthodes computationnelles plus élaborées, méthodes qui comme la TUA ou la TP,

¹ Voir Selten in Gigerenzer/Selten (Eds.) (2001), p 17.

² Comme le remarque Selten, *ibid.*, l'optimisation peut même conduire à une forme de régression à l'infini puisqu'au premier niveau il s'agit de déterminer l'alternative qui doit être choisie, mais pour ce faire il faut déterminer la méthode qui permettra de trouver l'alternative en question, etc. Par ailleurs l'utilisation de l'optimisation sous contrainte (de temps, d'argent, de ressource) peut avoir un coût très élevé puisque le sujet doit justement calculer à chaque étape ce que lui coûte la poursuite de la recherche d'information en plus de ce que lui coûte la décision en elle-même, il doit s'arrêter quand le coût de la recherche d'information dépasse le bénéfice de l'information recherchée, mais là encore nous sommes soumis au même problème : comment calculer ce coût ? La situation est la suivante : le calcul du coût de l'information implique un coût supplémentaire, voir Gigerenzer/Todd in Gigerenzer/Todd (Eds.) (1999), p 11. Le but des recherches de Gigerenzer et Al. est de montrer, dans la droite ligne de Simon (1955), que la collecte d'information doit être la plus rapide et la plus économique possible.

³ Comme le remarquent Gigerenzer/Todd in Gigerenzer/Todd (eds) (1999), p 6 : « Our premise is that much of human reasoning and decision making can be modeled by fast and frugal heuristics that make inferences with limited time and knowledge ».

⁴ Basées notamment sur le respect des axiomes du calcul des probabilités (Ramsey/Savage) et d'une axiomatique (NM). Mais nous avons vu comment cette prétention normative avait été sévèrement critiquée sans toutefois avoir été totalement dépassée.

⁵ La construction de l'environnement par le sujet est une idée implicite des articles de Simon (1955), (1956) et (1990) : *l'environnement c'est ce qui est pertinent pour la décision* et cette pertinence est construite par élimination.

⁶ L'idée fut d'abord exposée dans l'article fondateur de 1996 : « Reasoning the Fast and Frugal way : Models of Bounded Rationality », *Psychological Review*, 103, 650-669. Nous nous référerons à cet article dans l'édition de Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 31-58, cependant nous fonderons aussi nos analyses sur l'article de 2009 : « Homo Euristicus : why biased minds make better inferences », *Topics in Cognitive Science*, 1, 107-143, qui ouvre le volume précité. Cet article a le mérite de reprendre et de raffiner les thèses de 1996, notamment en y intégrant les résultats intermédiaires obtenus dans Gigerenzer/Todd (Eds.) (1999), ainsi que dans Gigerenzer/Selten (Eds.) (2001). Par ailleurs Gigerenzer a rendu accessible l'ensemble de la théorie de la rationalité limitée dans un livre de vulgarisation très bien fait : *Rationality for Mortals*. (2008).

supposent une pondération et une computation¹ à partir de toute l'information disponible² : « Here is the first important discovery : Heuristics can lead to more accurate inferences than strategies that use more information and computation »³

Aussi va-t-il falloir se poser certaines questions : comment le concept d'heuristique doit-il être repensé ? Quels types d'heuristiques vont devoir être convoquées ? Quelles sont les heuristiques les plus efficaces en contexte décisionnel ? Ces interrogations nous donneront des éléments pour penser, dans notre dernier chapitre, le travail heuristique du SRD. La découverte majeure que vont exposer Gigerenzer et Al. est la suivante : l'utilisation spontanée par les sujets d'un certain type d'heuristique – qui ne prend pas en compte la totalité de l'information disponible et qui ne cherche pas en acquérir plus – produit des résultats et des estimations plus précises que s'ils utilisaient, pour le même problème et dans le même environnement, des théories impliquant l'utilisation de la totalité de l'information. Ce fait est baptisé par les auteurs : « Less-is-more effect ». Cela permet de dépasser – outre les problèmes liés à la maximisation dans le TUA et la TP – la vision que proposait par exemple Stigler (1961) ou Anderson (1991) de l'optimisation sous contrainte où le problème de la collecte de l'information est relié à la dynamique entre précision de l'information et effort pour la produire (« Accuracy-effort trade off »). Théorie qui permettait d'expliquer l'utilisation des heuristiques – en quelque sorte par défaut – parce que justement elles sont moins exigeantes en temps et en effort, mais le problème c'est qu'ici aussi l'utilisation de ces « méthodes » était à verser au registre d'une certaine défaillance, ou tout au moins d'une certaine paresse intellectuelle.

¹ Voir supra, p 167 pour la définition du concept d'heuristique en opposition à la computation et au calcul. Définition d'ailleurs inspirée de Gigerenzer/Brighton (2009) in Gigerenzer/Hertwig/Pachur (Eds.) (2011). Pour Gigerenzer et son équipe il s'agissait au début de confronter la méthode de régression linéaire, utilisée notamment en sociologie, et en économie avec une méthode plus approximative que les auteurs baptisent « tallying » afin de montrer que bien que la seconde méthode ne pondère pas l'information disponible, elle fonctionne mieux que la régression linéaire dans certains contextes. Il fallait donc analyser le fait que « tallying can often match and even outperform complex calculation is relevant to understanding the nature of adaptive cognition », Gigerenzer/Brighton, p 6. C'est à partir de cette idée que les auteurs vont formuler le fait qu'avec moins de moyens cognitifs mis en œuvre on peut arriver à plus de résultats dans le domaine du choix rationnel.

² C'est là en effet un des problèmes majeurs des théories que nous avons étudiés dans notre premier chapitre et que nous retrouvons dans la TP : toutes supposent que le sujet a face à lui toute l'information disponible et qu'il doit seulement calculer pour trouver la solution optimale. Il y a un caractère nettement leibnizien dans cette prétention puisqu'elle suppose une calculabilité totale et la possibilité de rendre absolument transparent l'environnement informatif. C'est aussi une des faiblesses de la TP, puisque même si elle parvient à formaliser la distorsion de l'information probabiliste par le sujet, ce qui est un net progrès, elle ne parvient pas à sortir de cet horizon de calculabilité. Il conviendra de montrer, mais pour cela il faudra aussi aller plus loin que Gigerenzer, qu'outre le fait que l'information n'est jamais que partielle et fragmentaire, il y a souvent une résistance de l'environnement décisionnel lui-même.

³ Gigerenzer/Brighton (2009), p 5.

L'idée de Gigerenzer et Al. est la suivante : en contexte décisionnel un surplus d'information¹ ou de computation peut diminuer la précision, de ce fait, le décideur, ou, de façon équivalente dans le cadre de nos hypothèses, le SRD, utilise les heuristiques les plus simples possibles, les plus économiques quant aux moyens – basées sur la reconnaissance – afin d'être plus précis que s'il faisait appel à des méthodes plus élaborées. C'est un renversement de perspective complet par rapport à ce que nous avons vu jusqu'à maintenant : « The mind can use less information and computation or take less time and nevertheless achieve better performance »².

Il convient maintenant de se demander comment ce mécanisme fonctionne³. C'est l'analyse de la relation du sujet à son environnement⁴ qui va nous permettre de comprendre ce phénomène, car il faut garder à l'esprit que les thèses de Gigerenzer et Al. ont pour but premier d'expliquer comment un sujet⁵ peut et doit utiliser l'environnement de la façon la plus efficace possible. En effet une des caractéristiques essentielles de l'environnement, dans ce paradigme, est qu'il est constitué d'indices (« cues ») et de critères⁶ avec lesquels le sujet va interagir de façon spontanée. Le travail devra donc être orienté par le questionnement suivant :

« ...what statistical properties of the environment allow particular heuristics to perform well. Their identification enables us to predict in which environments a heuristics is used. We might then go on to ask whether such statistical properties are easy to

¹ En effet comme le remarquent Gigerenzer in Gigerenzer/Selten (Eds.) (2001), p 47-48, l'information peut tout simplement surdéterminer (« over fitting ») le problème ou la situation qui se présente. Par exemple la recherche d'indices dans un contexte décisionnel donné peut produire une quantité d'information trop importante, ce qui va induire une perte de temps : quel indice est pertinent ?

² Gigerenzer/Brighton (2009), p 9.

³ Il est bien entendu que l'utilisation de l'effet « moins c'est plus » devra être limité, de la même façon que l'utilisation des heuristiques par le SRD : son fonctionnement n'est pas exclusivement heuristique. Cependant nous pouvons d'ores et déjà faire remarquer que dans le cadre de la décision médicale en contexte d'urgence cette approche donne des résultats assez prometteurs, voir à ce sujet Gigerenzer/Brighton (2009), p 35-37 ; Gigerenzer/Todd in Gigerenzer/Todd (Eds.) (1999), p 4. Le contexte médical a d'ailleurs été largement investi par Gigerenzer/Craig (2011), non seulement dans le domaine de la décision, mais aussi dans le cadre de la transmission de l'information aux patients.

⁴ Il nous faut ajouter ici une précision lexicographique importante : à partir de maintenant lorsque nous parlerons d'« environnement » sans qualificatif il s'agira de l'ensemble des éléments qui entourent le sujet de façon exogène, lorsque nous parlerons d'« environnement décisionnel » il s'agira de la partie de la réalité que le sujet découpe par l'opération de visée.

⁵ En fait il faudrait peut-être généraliser l'utilisation de ces heuristiques à tous les organismes vivants, ou du moins l'utilisation de certaines règles empiriques constantes et adaptatives, par exemple dans le cadre de la recherche de nourriture ou de partenaires sexuels. Gigerenzer et Al. établissent de façon constante le lien entre les sujets humains et les organismes vivants, car il ne faut pas oublier que pour eux, dans la lignée de Tooby et Cosmides, le caractère adaptatif des heuristiques est central, voir à ce sujet Gigerenzer/Brighton (2009), p 2-3 ; mais surtout Hutchinson/Gigerenzer (2005) in Gigerenzer/Hertwig/Pachur (Eds.) (2011) p 110-133, article qui explore justement jusqu'où et comment le parallèle avec l'utilisation de règles empiriques par les organismes vivants doit être poussé.

⁶ Voir, Hutchinson/Gigerenzer (2005), p 110.

recognize, and hence how a heuristic for selecting appropriate heuristics might work »¹.

La réponse à cette interrogation ne sera donc pas à chercher seulement dans le travail de l'heuristique, mais dans le fait que telle heuristique ne peut fonctionner que dans tel environnement particulier. C'est l'essence même de la rationalité écologique². Pour nous il s'agit là d'une opération essentielle du système de référence décisionnel avec celle qui consiste à sélectionner à partir de son propre fonds la règle qui semble convenir au cas dans des situations inhabituelles³. Cependant cet extrait appelle encore au moins un commentaire : qu'entend-on de façon générale par environnement et par propriété statistique de l'environnement ?

Dans la lignée de Simon (1956), un environnement peut être défini de façon négative ainsi : « The term environment, here, does not refer to a description of the total physical and biological environment, but only to that part important to an organism, given its needs and goals »⁴. Transposée dans le domaine décisionnel cette définition nous indique qu'il s'agit donc ici de restreindre le concept d'environnement à ce qui est *utilisable par le sujet*, de façon la plus rapide et la plus économique possible. L'environnement est défini par sa capacité informative et sa pertinence contextuelle, non pas comme cette partie dynamique du monde dont l'ego constitue – sur le mode du *ici* – le point d'où partent toutes les perspectives⁵.

Gigerenzer propose donc une définition opératoire qui permet de conceptualiser la relation du sujet à son environnement⁶ : environnement qui est déterminé par des indices statistiques entre lesquels le sujet doit opérer une discrimination efficace. La question va être maintenant de déterminer quels mécanismes sont à l'œuvre dans ce processus de discrimination, c'est-à-dire quelles heuristiques sont utilisées par le sujet pour tirer

¹ Ibid., p 125.

² Voir Gigerenzer/Todd in Gigerenzer/Todd (Eds) (1999), p 5 ; Gigerenzer in Gigerenzer/Selten (eds) (2001), p 46-47.

³ Comme nous le verrons certaines décisions peuvent obliger les sujets à produire des règles inédites, de leurs propres fonds.

⁴ Gigerenzer in Gigerenzer/Selten (Eds.) (2001), p 39.

⁵ Voir Husserl (1952), p 223. Par la suite, et justement pour construire les concepts d'environnement et de résistance nous allons avoir recours à un certain nombre de ressources phénoménologiques, et en particulier du deuxième volume des *Ideen*.

⁶ Cependant ce qui est ici gagné de façon incontestable du point de vue du réalisme, à savoir l'intégration du découpage de l'environnement aux processus décisionnels, ne permet pas encore de constituer le concept de décision et de volonté de façon philosophique. En effet, à aucun moment Gigerenzer et Al. ne proposent une enquête philosophique sur la structure du SRD qui est la condition sine qua non de la constitution des concepts de décision et de volonté. C'est ce que nous nous proposerons de faire dans notre prochaine partie.

l'information nécessaire et suffisante de l'environnement¹.

Il va s'agir de faire appel à une connaissance limitée par l'environnement et par les capacités subjectives pour produire des inférences rapides, Gigerenzer n'hésite d'ailleurs pas à utiliser le terme de « guess » pour décrire le processus : deviner des éléments du monde que nous ne connaissons pas à partir d'indicateurs incertains. L'exemple sur lequel Gigerenzer et Al. ont basé leurs recherches originales, et qui a donné lieu à plusieurs expériences est le suivant : on demandait à des étudiants américains et allemands d'évaluer la population d'un certain nombre de villes allemandes en prenant en considération l'environnement suivant : l'ensemble de toutes les villes allemandes de plus de 100 000 habitants avec une série de neuf indices d'une validité écologique et d'un taux discriminatoire différent². La question était : « d'après vous, la ville X possède-t-elle ou non plus de 100 000 habitants ? » Il apparut que ce furent les étudiants américains qui obtinrent les meilleurs résultats quoique bénéficiant d'une proximité géographique moindre, et par définition d'une « connaissance » plus vague. On fit l'expérience dans le sens inverse, et le résultat fut aussi inversé, pour les villes américaines ce furent cette fois les Allemands qui eurent les meilleurs résultats³.

Les sujets utilisent donc un algorithme⁴, économique et rapide qui pour être le plus efficace possible doit remplir certaines conditions. Il s'agit ici de décrire le contenu et le fonctionnement de ce que Gigerenzer et Al. appellent la « boîte à outils adaptative »

¹ Les analyses qui vont suivre se situent dans le cadre de ce que Gigerenzer et Al. appellent la théorie des modèles mentaux probabilistes (« probabilistic mental models »), dont l'idée centrale est la suivante : les inférences produites par les sujets concernant les états du monde inconnus pour eux reposent sur des indices probabilistes tirés de l'environnement qui permettent de mettre en œuvre l'heuristique centrale de la théorie : *la recognition*. Le schéma est le suivant : les indices causent ou non la recognition, il y a donc une valeur discriminatoire des indices. La recognition produit une connaissance qui servira de fondement pour la décision. Voir, Gigerenzer (1993), et Gigerenzer/Goldstein (1996). *Les probabilités sont utilisées ici pour désigner des fréquences d'événements* dans une classe de référence donnée. Par exemple de ce point de vue là il est fréquent qu'une ville possédant plusieurs universités soit plus importante qu'une ville qui n'en possède pas, même chose avec une équipe de football internationale. Le nombre d'universités peut donc servir d'indice *fréquentiste* discriminatoire dans l'optique où nous voudrions déterminer si telle ou telle ville est ou non une grande ville.

² Pour plus de détails, voir Gigerenzer/Goldstein (1996) in Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 39-41. En plus de ceux cités dans la note précédente il y avait aussi : la ville est-elle une capitale nationale ? Fait-elle partie de la ceinture industrielle ?

³ L'ensemble des résultats fut comparé avec d'autres résultats qui avaient été obtenus par des méthodes plus élaborées comme la régression linéaire : « Amazingly, the predictive accuracy of this heuristic [take-the-best], judged on a real-world data set about German cities, was about equal to, or better than, that of multiple regression », Hutchinson/Gigerenzer (2005) in Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 111.

⁴ Voir Gigerenzer/Goldstein (1996). L'utilisation du terme « algorithme » peut sembler ici paradoxale d'autant plus qu'on a coutume de l'opposer au concept d'heuristique. Si Gigerenzer et Al. se permettent cette utilisation c'est en raison de la pertinence de la méthode utilisée qui joue le rôle normal d'un algorithme : elle permet tout simplement de répondre à la question en un nombre fini d'étapes. L'algorithme utilisé par les sujets ici est baptisé par les auteurs : « take-the-best » dont le but est le classement subjectif des indices par ordre de validité. Dans cet algorithme, la validité d'un indice peut être calculée en rapportant le nombre d'inférences correctes établies à l'aide de cet indice au nombre d'inférence qu'il est possible d'établir en utilisant le même indice, voir Gigerenzer/Brighton (2009).

(« adaptative toolbox ») ¹ qui doit contenir comme structures de bases trois règles simples concernant l'environnement et les indices² :

- Une règle de recherche qui doit permettre de trier les indices en fonction de leur validité.
- Une deuxième règle de recherche qui doit permettre de discriminer les indices en fonction de leur accessibilité environnementale.
- Enfin une règle d'arrêt. Stopper les recherches à partir du moment où deux indices pointent vers le même objet, et inférer de là que cet objet a la plus haute valeur en ce qui concerne les critères en question.

C'est ainsi que fonctionne la cognition des sujets en contexte de recherche d'information, tout le processus doit être le plus simple et le plus efficace possible afin de tirer le parti le plus rapide et le plus optimum de l'information disponible. Le fait d'insérer le rapport du sujet à l'environnement dans le processus de prise de décision permet d'ajouter une dimension réaliste non négligeable, sans pour autant permettre de construire le concept philosophiquement.

Nous allons maintenant nous tourner vers une heuristique centrale pour Gigerenzer et Al. : la recognition³, qui permet de conceptualiser deux choses : comment les sujets utilisent l'information qu'ils ont en mémoire, et comment la connaissance tacite que les sujets possèdent peut être actualisée au moyen d'indices environnementaux. La recognition ne doit pas être confondue avec l'heuristique de disponibilité contre laquelle K/T nous mettaient en garde dans leur article de 1974⁴. À la différence de la disponibilité qui s'applique à des éléments mémorisés et qui est mesurée par la vitesse ou l'ordre dans lequel ils sont remémorés, la recognition concerne la différence entre des éléments qui peuvent être dans la mémoire du sujet mais aussi à l'extérieur. La recognition est une forme cognitive de ressouvenir à caractère épistémique⁵.

¹ Gigerenzer/Todd in Gigerenzer/Todd (Eds.) (1999), p 30, la définissent ainsi : « the collection of specialized cognitive mechanisms that evolution has built into the human mind for specific domains of inference and reasoning ».

² Il faut noter que les indices ne viennent pas tous de l'environnement, certains peuvent faire l'objet de recognition et sont donc ainsi à relier avec la mémoire. Nous allons analyser sous peu l'heuristique de recognition.

³ Ce concept est analysé en profondeur dans Goldstein/Gigerenzer (2002) in Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 60-81.

⁴ Voir Tversky/Kahneman (1974), in Kahneman/Slovic/Tversky (Eds.) (1982), p 4-11, et notre partie sur cette question.

⁵ Nous employons le terme dans son sens large, la connaissance étant entendue dans son sens non technique, incluant les opinions, etc.

La recognition est définie ainsi par Gigerenzer/Goldstein (2002) : « *Recognition heuristic* : if one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion »¹. La pertinence de cette heuristique – qui est présentée comme une règle d’inférence – doit toujours être ramenée à la structure de l’environnement, c’est-à-dire qu’elle ne fonctionne pas dans tous les cas possibles. Quand nous parlons de la structure de l’environnement nous faisons référence au problème décisionnel particulier que le sujet doit résoudre et dont la délimitation est la première tâche du SRD. Comme nous le remarquons plus haut *l’environnement est toujours relatif et défini par le problème posé*². Par exemple la recognition ne fonctionnera pas dans un environnement où l’information surdétermine le problème, car la caractéristique essentielle de cette heuristique c’est qu’elle est basée sur un *manque* d’information plus ou moins important des sujets en fonction du problème posé³. C’est aussi cette non-universalité, et cette dépendance par rapport à l’environnement qui fait toute la différence entre la théorie de Gigerenzer et Al. et la TUA ou la TP.

Nous avons donc à faire ici à un renversement de perspective assez important concernant la définition de la rationalité en contexte décisionnel : il ne s’agit plus d’optimiser ou de maximiser une fonction. D’ailleurs la recognition peut même fonctionner avec des informations de « seconde main ». C’est ce que conceptualise le rôle des *médiateurs* que nous allons analyser rapidement. Ce concept est convoqué afin de thématiser le fait que l’information peut être obtenue de façon indirecte : « There are ‘mediators’ in the environment that have the dual property of reflecting (but not revealing) the criterion and being accessible to the senses »⁴. Dans le cadre de la théorie de Gigerenzer et Al. un *critère* est un élément du contexte décisionnel qui peut être considéré comme un indice ou comme un élément permettant d’opérer une discrimination, ces critères peuvent être accessibles de façon directe, comme c’était le cas pour l’exemple des villes allemandes, ou pas. Ces critères contiennent en

¹ Goldstein/Gigerenzer (2002), p 61.

² Dans notre prochaine partie nous verrons que cette thèse doit être nuancée, il y a des éléments plus complexes et divers qui rentrent en jeu, et il semblerait que Gigerenzer et Al. en aient eu connaissance lorsqu’ils parlent métaphoriquement d’environnements bruyants, c’est-à-dire d’environnement parasité par d’autres informations que celles pertinentes pour le problème.

³ La validité de cette heuristique est formalisable, elle est équivalente à : $R/R+W$, où R est le nombre d’inférences correctes que la recognition peut atteindre, R est calculé à partir de toutes les paires d’objets où l’heuristique est parvenue à en discriminer un, W est le nombre d’inférences incorrectes sous les mêmes conditions.

⁴ Goldstein/Gigerenzer (2002) in Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 63.

eux une information de type statistique explicite ou implicite. Par exemple, imaginons un individu qui désire placer un de ses enfants dans une des meilleures universités de son pays. Comme il n'a pas idée de la dotation réelle des universités, il pourra se rapporter dans son choix au nombre de fois qu'une université X est citée dans le journal. Cela lui donne une information sur l'importance de l'établissement, mais elle est indirecte. Le journal joue ici le rôle du médiateur entre le critère (la dotation) et la reconnaissance.

Le lien entre le critère et le médiateur est défini comme une corrélation écologique, dans le sens où elle permet de tirer parti de la structure de l'environnement. Dans notre exemple, une corrélation de substitution s'opère entre le médiateur et le contenu de la reconnaissance¹, contenu qui peut être par exemple une liste de noms d'universités entendus çà et là puis mémorisés, contenu qui va être corrélé avec le nombre d'instances où le journal parle de l'université X. Le processus permettra alors d'opérer une corrélation de reconnaissance entre le critère (dotation de l'université) et le contenu de la reconnaissance. Ce schéma est baptisé par les auteurs « la rationalité écologique de l'heuristique de reconnaissance »², il permet de construire la décision sur la base d'une connaissance incertaine et même parfois indirecte.

Ce qui nous intéresse dans la théorie de Gigerenzer, c'est d'une part qu'elle va nous permettre dans notre troisième chapitre de penser le fonctionnement heuristique du SRD, d'autre part, elle nous permet de mettre le rapport de l'individu à l'environnement au centre du processus décisionnel. Cela constitue un pas important dans notre travail, puisque nous allons, dans les deux parties suivantes montrer comment le SRD construit l'environnement décisionnel, nous tenterons aussi de restituer toute la densité qu'il a dans la réalité. D'ailleurs les médiateurs que nous venons d'étudier nous mettent sur une piste quand à la constitution de cet environnement informatif : il est orienté. Premièrement, parce que les éléments de l'environnement sont interconnectés et reliés entre eux. L'environnement est construit et orienté par les besoins du sujet en contexte décisionnel, ce que nous appellerons sa « visée ». Une décision est toujours un certain regard et par là même une certaine orientation qui s'inscrit dans et contre le monde. Cependant avant de constituer l'environnement décisionnel comme objet philosophique, il s'agira de comprendre dans quelle mesure nous devons ou pas nous détacher du concept classique de contexte décisionnel.

¹ Ibid., p 63, pour le schéma.

² Ibid., p 64.

2) Du contexte à l'environnement.

a – **Insuffisance du concept de contexte**¹.

Les travaux de Gigerenzer et Al. nous ont permis de nous faire une première idée de l'importance du lien dynamique qui unit le sujet à l'environnement. Environnement qui est lui-même, dans la théorie de Gigerenzer, totalement déterminé par l'information pertinente pour le sujet en contexte de décision². Il va s'agir maintenant de penser le passage du concept de *contexte probabiliste* à celui d'*environnement décisionnel*. Comme nous l'avons vu dans le chapitre précédent en étudiant l'article d'Ellsberg (1961), il existe à l'intérieur du paradigme bayésien et du modèle N/M des problèmes liés au contexte probabiliste qui donnent lieu à des paradoxes, par exemple du fait de l'aversion des sujets pour l'ambiguïté.

Si nous prenons en compte les analyses d'Ellsberg (1961), de K/T (1979)³, et Gigerenzer⁴ et Al. nous nous apercevons que – dans le cadre d'une analyse philosophique et non plus strictement mathématique⁵ – le concept de contexte tel qu'utilisé dans le bayésianisme et la TUA doit être modifié, repensé, voire dépassé. Et ce dépassement ne peut s'opérer que vers un concept qui tient compte de tous les éléments représentationnels qui interviennent dans la décision, il va donc s'agir de penser le concept d'environnement en connexion avec le sujet du point de vue de la décision, plus précisément comme ce que le sujet découpe et délimite de la réalité en vue de la décision. Le concept d'environnement

¹ À partir de maintenant, le mode d'analyse, d'exposition et de composition va être légèrement modifié : nous allons passer d'un registre qui était basé sur l'analyse de textes et d'articles, à un registre plus conceptuel et pour tout dire plus dissertatif. Pour cela nous procéderons à une légère mise à distance des textes dans leur littéralité, et ce afin de pouvoir donner à nos hypothèses la place et l'extension qui leur conviennent. Mais nous ne tournons pas le dos à nos analyses précédentes qui, bien au contraire, feront l'objet d'une reprise synthétique et sélective dont le but sera de les replacer dans une perspective philosophique. Pour plus de détails sur ce changement méthodologique et sur le passage à une analyse plus conceptuelle, voir nos remarques dans l'introduction générale.

² La suite des analyses va tenter de montrer que l'environnement décisionnel est une entité beaucoup plus riche et beaucoup plus complexe. Il y a toutefois beaucoup d'éléments importants pour la philosophie à tirer des recherches des Gigerenzer et Al. sur les heuristiques et leurs liens avec le concept d'environnement, voir par exemple les articles réunis sous le titre « Everyday things » aux pages 712-759 de Gigerenzer/Hertwig/Pachur (Eds.) (2011).

³ Rappelons que les problèmes de pondération des probabilités avaient été pressentis entre autres par Allais (1953), Edwards (1954) et (1961).

⁴ Les travaux de Gigerenzer ne rentrent pas à proprement parler dans la théorie classique de la décision, cependant l'auteur a alimenté le débat de façon critique, dans deux articles de 1991 et 1996. L'auteur, qui est un porte-parole de l'interprétation fréquentiste en théorie des probabilités, avait montré que certains biais présentés dans Tversky/Kahneman (1974) peuvent être évités si nous nous basons sur des fréquences d'événements et non sur les probabilités des événements particuliers. D'autre part, il existe un article de Katsikopoulos/Gigerenzer in Gigerenzer/Hertwig/Pachur (Eds.) (2011), p 186-200, dont le but est de montrer que grâce à l'« heuristique de priorité » (« priority heuristic ») on peut expliquer simultanément l'ensemble des violations de la TUA. Il faut noter que Gigerenzer a toujours été très critique en ce qui concerne la portée explicative de la TP et de la TCP de 1992, pour lui les deux théories ne permettaient pas d'expliquer *toutes* les violations de la TUA.

⁵ Il est important de relever ce fait car dans le cadre de théories formelles, le concept de contexte probabiliste reste opératoire.

décisionnel sera construit afin de conceptualiser la totalité de l'interaction du sujet – par le biais de la représentation – avec le monde dans le cadre du choix rationnel. Il faudra donc par exemple penser et analyser les concepts d'*orientation* et de *résistance* tout simplement parce qu'il n'y a pas de décision qui ne soit par exemple temporalisée¹, aussi rapide soit-elle, d'autre part il n'y a pas de décision qui ne se fasse sur fond d'une certaine résistance.

Avant de passer à la construction philosophique du concept d'environnement nous allons revenir sur la définition classique du contexte décisionnel et sur ses insuffisances. Le premier élément qui limite la portée de l'utilisation de ce concept, c'est qu'il a été thématiqué afin de rendre compte d'une information de type mathématique : la probabilité². Or, la probabilité peut soit être une donnée objective de l'option, ou de la perspective comme c'est le cas chez N/M³, Samuelson, Marschak, etc., soit la mesure d'un degré de croyance comme c'est le cas chez Ramsey (1926) et Savage (1972). Cependant dans les deux cas de figure nous pensons que cette mesure probabiliste échoue à rendre compte de la complexité de l'environnement décisionnel, c'est ce que mettent en avant les critiques qui émergent avec Allais (1953) et se continuent avec Ellsberg (1961). Pour le premier la TUA néglige la perception des distributions de probabilités par le sujet : les formalisations de N/M et Savage manquent tout simplement l'impact que la *représentation* – en particulier la probabilité de perdre – a sur les décideurs. Pour Ellsberg, comme nous l'avons vu plus haut, la distinction classique des contextes sous détermine la réalité décisionnelle. Et c'est cette sous-détermination qui produit les paradoxes que nous avons étudiés, paradoxes qui ne peuvent être évités qu'en ajoutant, comme le fait la TP, en distinguant deux phases dans la prise de décision et en introduisant des termes pondérateurs dans les équations, termes qui rapprochent ainsi les modèles de la réalité.

Un autre aspect qui a été évoqué sans être jusqu'ici discuté, mais qui mérite l'attention est le suivant : dans quelle mesure les probabilités sont-elles un outil fiable pour mesurer le degré de croyance ? C'est une des thèses centrales du bayésianisme⁴. Là aussi il convient de nuancer les choses et d'affiner la question en nous demandant si toutes les croyances peuvent être mesurées par les probabilités. Nous faisons ici l'hypothèse que, de la même façon que

¹ Comme nous le verrons, le temps est un élément central de toute décision.

² Pour rappel, voir Luce/Raiffa (1957), p 12-23 ; Saint-Sernin (1973), p 19-22; Peterson (2009), p 5-8.

³ Voir N/M (1947), p 19.

⁴ Voir Bradley (2006) ; Chater/Oaksford (2008), p 13-20 ; Manktelow (2012), p 246.

certaines décisions échappent à la formalisation de la TUA ou du bayésianisme, certaines croyances se prêtent mal à une mesure probabiliste, et ce pour la raison qu'elles font partie d'un système dans lequel la valeur d'une croyance est fonction de toutes les autres. De la même manière que dans un holisme épistémologique¹ on ne teste pas une hypothèse singulière, mais un système d'hypothèses, dans le cadre du holisme de la croyance que nous soutenons ici, on ne mesure jamais une croyance isolée, mais un système de croyances subjectif qui est une partie centrale du système de référence général et auquel est connecté le SRD².

Enfin, dernier point, si nous partons de l'idée que les contextes probabilistes fonctionnent dans la réalité, cela signifie que – du fait de la possibilité de calculer – tout choix a une solution optimale. Or rien ne nous confirme que les choses se passent ainsi : il y a bon nombre de décisions qui n'ont pas d'optimum³, ce qui ne veut pas dire qu'aucune n'en a. Il ne s'agit plus ici de critiquer le temps et l'effort nécessaires au sujet pour maximiser mais de constater que certaines décisions ne rentrent tout simplement pas dans ce cadre-là.

Ce que met en avant le concept de contexte, même avec ces insuffisances, c'est que la décision est un processus cognitif qui fait fond sur une extériorité que le sujet se représente et qu'on ne peut ignorer. Dans le cadre de la TUA l'information probabiliste et la valeur des résultats étaient convoquées pour formaliser le minimum de cette extériorité, or nous faisons l'hypothèse que cette information – quand elle existe et quand elle est exploitable par le sujet – ne représente qu'un élément d'un ensemble beaucoup plus vaste, complexe et ramifié : l'environnement décisionnel, qui, – comme l'ont montré Simon et Gigerenzer – est cette partie de l'extériorité que le sujet découpe, sélectionne, ou délimite lorsqu'il prend une décision, et dont la caractéristique première est d'être *orienté*, par exemple selon le temps et selon l'espace, orientation qui nous indique une possible hiérarchisation des environnements décisionnels selon la proximité ou la distance⁴ par exemple, ou selon les concepts d'immédiateté ou de délai.

¹ Voir Gillies (1998).

² Voir notre schéma, p 219.

³ Voir Hutchinson/Gigerenzer (2005), p 124.

⁴ Ici nous pouvons penser par exemple aux environnements quotidiens qui sont définis par une certaine « manipulabilité » une certaine proximité des lieux, une certaine disponibilité des objets et des ustensiles connus, il s'agit là de l'habitat, de la maison, du chez-soi, qui constitue à lui seul tout un environnement stable dans lequel le décider opère de façon particulière. Voir par exemple, Heidegger (1927), p 84-88, Levinas (1971), p 61-162, Romano (2010), p 673-726. Nous reviendrons sur ce problème, notons seulement que dans notre thèse nous soutenons aussi l'idée que plus un environnement est stable,

Avant de passer à l'analyse du concept d'environnement décisionnel nous allons nous poser deux dernières questions concernant le contexte probabiliste en théorie classique de la décision : les probabilités doivent-elles être abandonnées totalement comme composantes de l'environnement décisionnel? Qu'est-ce qui précisément rend le concept de contexte difficilement utilisable du point de vue philosophique dans notre travail ?

Encore une fois, gardons-nous de donner des éléments de réponse trop tranchés, et nuançons encore un peu. L'hypothèse que nous faisons ici est que dans la plupart des cas les probabilités ne peuvent plus jouer le premier rôle avec la valeur des résultats¹ dans la prise de décision. Comme nous l'avons remarqué, les sujets ont une vision assez floue des probabilités², en tout cas une vision qui dans beaucoup de cas ne correspond pas du tout aux probabilités axiomatisées, et cette conception ne semble pas – comme le soulignait Gigerenzer – priver les décideurs d'une cohérence dans un large panel de situations. Cela nous indique, contre les idées de Ramsey et Savage, que les probabilités ne peuvent pas toujours constituer un code de conduite normatif.

Une autre question est la suivante : quelles sont les limitations philosophiques du concept de contexte en théorie de la décision ? Ici il convient de soulever trois points, dont deux recevront un traitement plus approfondi dans la suite de notre analyse. Tout d'abord, le concept de contexte probabiliste exclut toute autre forme d'information que celle qui porte sur les probabilités des résultats. Outre le fait que Ellsberg (1961) avait mis en avant que le contexte ainsi construit pouvait mener les sujets à produire certains choix paradoxaux, les recherches de Gigerenzer/Goldstein (2009) ont montré qu'une décision repose sur d'autres éléments informatifs et surtout sur la capacité qu'ont les sujets à les utiliser le plus rapidement et le plus efficacement possible. Le concept de contexte est donc trop réducteur, il échoue à rendre compte de la complexité du réel³, il y a toute une strate de l'information qui lui échappe, par exemple l'influence que peut avoir la présence ou le regard d'autrui sur nos

maîtrisable, contrôlable, et plus il se rapproche d'un « petit monde » à la façon dont le définit Savage (1954), p 83-84, et la question qu'il faut se poser alors est celle-ci : à partir du moment où nous sommes dans ce type d'environnements peut-on appliquer les axiomes de la TUA. Nous pensons que la réponse est positive.

¹ Puisque la plupart du temps la valeur des résultats n'est tout bonnement pas quantifiable. À moins d'opérer une simplification qui enlève du même coup tout intérêt à la question.

² C'est ce qui ressort des analyses du chapitre deux.

³ Complexité que Merleau-Ponty (1945) exprime ainsi : « Nous sommes mêlés au monde et aux autres dans une confusion inextricable », p 519.

choix¹, lui échappe aussi tout ce qui fait l'épaisseur de la réalité vécue comme la dimension spatio-temporelle et orientée de tout environnement et de toute subjectivité.

C'est justement la seconde carence majeure du contexte probabiliste : mettre totalement de côté tout ce qui appartient à la subjectivité du décideur, et cette dimension est majeure pour la raison qu'elle est le centre de gravité du décider, et qu'elle permet de penser la singularité des décisions. En effet cette dimension est celle du SRG où prennent place les croyances et les valeurs qui sont à elles seules tout un ensemble de coefficients qui transforment et pondèrent l'environnement que le sujet se représente.

Enfin, dernier point de la critique : le contexte en théorie classique de la décision ne prend pas en compte l'aspect spatio-temporel² de la réalité en général et de l'environnement décisionnel en particulier. Par exemple l'absence de prise en compte du temps dans la TUA³ s'accommode mal d'une réalité dont une des caractéristiques majeures est d'être *dynamique*. Par ailleurs, de la même façon que la réalité est dans une *dynamique perpétuelle*, dont il faudra détailler les dimensions, le sujet lui-même est soumis à une *évolution* dont il faut tenir compte. Car si il est important de thématiser le rôle du système de croyances dans les décisions, il est tout aussi important de montrer que le SRG dont elles sont une partie est lui aussi soumis à une certaine dynamique, assez complexe d'ailleurs et qui provient de différentes formes d'interactions.

b – Le découpage de l'environnement décisionnel et son orientation : la visée.

Le découpage de l'*environnement décisionnel* à partir de la réalité extérieure et de la configuration interne du SRG est la première opération du SRD, il s'agit de ce que nous appellerons de façon synonyme : *établissement de la visée*. Nous allons ici garder un élément de la définition que donnent Simon et Gigerenzer de l'environnement décisionnel : sa

¹ Allais (1952), p 57-60, avait eu l'intuition de ce phénomène lorsqu'il expose les facteurs secondaires qui peuvent influencer une décision. Il nous semble que la prise en compte d'autrui comme élément déterminant de l'environnement décisionnel est essentielle. L'altérité, qu'elle soit familiale, intime, amicale ou sociale ne peut être exclue de la sphère de la décision. C'est ce que montrent par exemple les analyses de Husserl (1952), p 267-279, d'une façon plus générale on peut se reporter aux textes sur l'intersubjectivité traduits et édités par N. Depraz en 2001 ; voir aussi Merleau-Ponty (1945), p 405 ; et Levinas (1971), p 215-220.

² C'est un fait que N/M assument totalement, voir (1947), p 19.

³ Il y a eu cependant des tentatives pour formaliser la décision comme processus dynamique dans le cadre de la TUA, voir par exemple McClennen (1990), et, d'une façon générale le bayésianisme, par le biais de la formalisation de la révision des croyances est aussi une théorie dynamique, voir Edwards (1961) ; Savage (1954), p 46-50. Cependant, encore une fois il s'agit là d'une prise en compte du temps qui fait abstraction du temps perçu par le sujet en contexte décisionnel, ces théories ne peuvent par exemple rien nous dire du concept de délai.

caractéristique première est d’être délimité par la décision à prendre. La visée est construite par le sujet, et doit, de façon première être définie par le concept d’*orientation*, concept qui est contenu virtuellement dans les travaux de Simon (1956), (1990), et ceux de Gigerenzer/Brighton (2009). Seulement, comme nous allons le montrer cette orientation s’opère à partir de *plusieurs niveaux* et pas seulement à partir du concept de la recherche et de l’utilisation de l’information.

À un premier niveau, l’environnement décisionnel est orienté par la *subjectivité du décideur* – qui s’exprime dans la configuration du SRG – et son système de référence décisionnel, il s’agit de ce que nous pourrions appeler sa « visée » générale, visée qui est à la décision ce que l’intentionnalité est à la conscience d’une façon plus générale¹, puisque comme la conscience est toujours « conscience de », la visée est toujours « visée de » ; il n’y a jamais dans le SRD, lorsqu’il fonctionne, de vacuité quant au but. La visée n’est pas strictement équivalente au but² – elle le surdétermine légèrement –, elle pourrait être définie comme une *tension* qui s’opère entre le sujet et le monde extérieur, une attention particulière qui le délimite, qui circonscrit cet espace de la réalité comme le lieu possible du décider pour cette décision particulière, c’est-à-dire comme environnement décisionnel. C’est donc la visée qui constitue cet environnement-ci comme pertinent pour la décision.

Cette tension constitutive de la visée est à son tour conditionnée par le rapport du décideur à son corps propre³ et à la spatialisation de ce dernier qui donne une orientation à l’environnement immédiat⁴. Notre organisation psycho-perceptive⁵ donne à notre environnement corporel une orientation naturelle. Nous pourrions dire aussi que le corps propre est le premier objet de notre environnement⁶, et c’est un objet qui entre pour une bonne part dans nos décisions⁷ comme nous allons le voir sous peu en analysant comment il peut les favoriser ou les entraver⁸. Le corps comme lieu du sentir et des affects influence de façon

¹ Voir Husserl (1913), p 115-121.

² Elle permet en fait de déterminer le « lieu » dans lequel le but peut trouver à la fois son expression et son accomplissement. Voir à ce sujet la définition que donne Proust (2005), p 85, de l’intention et qui se rapproche beaucoup de ce que nous entendons ici par visée.

³ Nous empruntons ce concept à Merleau-Ponty (1945), p 127-183.

⁴ Voir Husserl (1952), p 227 et p 297.

⁵ Pour une analyse de la construction visuelle de notre environnement, voir Marr (1982), p 295-328.

⁶ Voir Husserl (1952), p 284 et 296. Merleau-Ponty (1945), p 406, dira même que le corps est le premier objet *culturel* de notre environnement.

⁷ Voir à ce sujet Schwarz (2002), in Gilovitch/Griffin/Kahneman (Eds.) (2002), p 545-546.

⁸ La problématique du corps comme « organe du vouloir » pour reprendre l’expression de Husserl avait totalement été mise de côté par la TUA et la TP, ce qui est une carence philosophique majeure. Le corps de par sa double nature de corps physique et

certaine aussi bien la constitution du système de référence décisionnel que son fonctionnement¹ ; par ailleurs, l'*ici* constitué par l'insularité de mon corps représente aussi bien l'origine de toute orientation décisionnelle possible² que la condition de possibilité de la subjectivité elle-même. *Le sujet décisionnel n'est sujet qu'en tant qu'il est situé dans un monde.*

Cette orientation originelle de la subjectivité nous amène à considérer deux niveaux d'orientation supplémentaires mais non moins essentiels de l'environnement décisionnel : le temps et l'espace qui sont des concepts fondamentaux, car ils sont connectés par exemple aux concepts de délai, de proximité et de distance. D'autre part, le temps et l'espace se déploient dans deux dimensions : à l'intérieur du sujet et dans la réalité extérieure. Nous aurons tôt fait de nous rendre compte que les deux formes de temporalités ne sont pas toujours synchrones³, et que la perception du temps peut-être distordue par un coefficient de pondération temporel, c'est ce qui se passe lorsque le sujet est confronté à l'attente ou au délai.

La TUA, outre le fait qu'elle ignore la totalité de la dimension subjective, ignore aussi le fait que les décisions peuvent être dynamiques en ce sens que le résultat de l'une peut conditionner celui de la suivante et être conditionnée par le résultat de celle qui précède⁴. Des situations décisionnelles de ce type se présentent dans l'existence, situations où les décisions se conditionnent les unes les autres un peu à la façon dont la conclusion d'un raisonnement peut devenir la prémisse du suivant. La TUA fait l'hypothèse que les résultats se situent dans un futur proche, que les décisions sont uniques et ne sont pas intriquées dans un tissu décisionnel et cognitif plus vaste. Dans le cas où les décisions sont concaténées les résultats sont reliés de deux façons et peuvent avoir deux effets sur l'environnement.

Dans le premier cas de figure, les décisions sont liées, mais seulement par le but général, par exemple lorsque le sujet va faire ses courses il a un but général qui passe par des

de corps ressenti, c'est-à-dire capable de souffrance, ne peut être exclu de la sphère du décider. Voir aussi Ricoeur (1950), p 117-177, et p 389-427.

¹ Husserl, op. cit., p 215-216.

² Husserl, op. cit., p 223.

³ Voir, par exemple Droit-Volet/Meck (2007), p 505 ; et pour une approche plus philosophique et phénoménologique du problème voir Jocelyn Benoist, in Jocelyn Benoist (2008) (Ed.), p 11-28. Ici nous pouvons dire que les résultats de la psychologie cognitive sont venus confirmer des intuitions sur le temps qui traversent la philosophie depuis le livre XI des *Confessions* jusqu'au débat entre Husserl et Bergson, voire même jusqu'à Jankélévitch.

⁴ Voir Edwards (1962), p 63-66. Edwards propose dans cet article sa propre taxinomie des environnements, taxinomie qui est fondée sur le concept de recherche de l'information en connexion avec trois types d'environnements stationnaires et deux types d'environnements non-stationnaires. Voir aussi McClennen (1990) ; Ariely/Zakay (2001), p 188-195. Comme le soulignent d'ailleurs les auteurs, il n'y a rien dans le domaine psychologique qui ne soit soumis au temps.

sous-buts (et donc des sous-décisions) : acheter tel ou tel item. Ici comme la visée est générale et homogène, l'environnement n'est pas modifié par les sous-décisions¹. Dans le deuxième cas les choses sont un peu différentes car la décision présente dépend partiellement ou totalement du résultat du choix passé, dans cette optique là l'environnement se précise, se détermine ou se transforme au fur et à mesure que le sujet progresse dans le temps. Le schéma est le suivant : la prise de décision produit l'élaboration d'une nouvelle visée, d'un nouveau but, ce qui entraîne à son tour une autre édition et une autre évaluation. Si nous nous situons dans le cadre de notre schéma page 219, sous l'item « inscription dans la réalité » vient se situer une autre visée, conditionnée par la réalisation de l'alternative².

Il reste encore deux caractéristiques, deux dimensions de l'environnement qu'il faut explorer avant de passer au concept de résistance : le fond sur lequel se détache l'information et autrui, c'est-à-dire la société, comme environnement. Les analyses de Todd (2001)³ mentionnent parfois la possibilité pour l'environnement d'être « bruyant » (« noisy »), c'est-à-dire que l'environnement même s'il est circonscrit par la visée du décideur peut ne pas être totalement transparent à la décision, et de ce fait la sélection de l'information pertinente peut être empêchée ou rendue difficile par des éléments qui ne sont pas des indices et viennent parasiter la situation qui doit alors être reconstruite par le sujet en fonction du but à atteindre afin d'éliminer les éléments perturbateurs.

Deux séries de remarques s'imposent dans le cadre de cette problématique. La non-transparence de l'environnement pour le décideur est à verser au mode d'apparition de la réalité elle-même : l'environnement peut être inclus d'une façon inextricable dans un système d'objets, d'êtres vivants, de signes et de signaux qui constituent le monde dans son mode d'apparaître sa complexité ainsi que sa quotidienneté. L'information ne se détache pas de

¹ En partant de l'hypothèse que nous nous situons dans le cours « normal » des choses. En effet si un incendie se déclare dans le supermarché, résultats et environnement changent de façon immédiate, ce qui va de soi. Remarquons aussi que dans le cadre de notre exemple les décisions peuvent être indépendantes ou pas, selon la catégorie d'objets considérés et leur homogénéité à l'intérieur de la catégorie elle-même. Par exemple le fait d'acheter du fromage peut susciter en moi l'envie d'acheter du vin et du pain. Nous avons ici un sous-ensemble, lié au but général et dont les sous-buts sont liés entre eux. Mais d'un autre côté le fait d'acheter du fromage ou pas n'a aucune influence sur le fait d'acheter de la lessive car mon baril est vide et que je dois le changer. Dans ce cadre-là, nous remarquons que le principe d'indépendance et le PCS fonctionnent parfaitement.

² Voir Gonzalez (2005), p 144.

³ Voir par exemple, Todd in Gigerenzer/Selten (Eds.) (2001), p 66.

l'ensemble de façon nécessairement claire, elle doit être abstraite de ce contexte plus large¹. Mais encore une fois il convient de nuancer les choses : plus l'environnement sera familier et moins il sera difficile pour le sujet de dissocier, de délimiter l'information pertinente sur le fond du monde.

Par exemple si le sujet est chez lui, qu'il décide de se préparer un café, l'environnement sera complètement transparent à sa décision, et toute l'information est disponible, même s'il n'y a plus de café : la visée travaille et délimite un microcosme pertinent qui est celui de l'habitat et de la manipulabilité. Il en ira tout autrement du même sujet décidant de partir s'installer à l'étranger, où pour le coup il va se trouver face à tout un ensemble informatif plus ou moins clair dans lequel il devra opérer un tri sélectif afin de *constituer un environnement décisionnel pertinent*.

Deuxième remarque, liée à la première : le monde est un système d'objets et d'êtres orientés les uns vers les autres, les uns en fonction des autres, et les uns pour les autres², il apparaît sous une forme holistique et il peut donc être difficile de dissocier ce qui est pertinent comme information de ce qui ne l'est pas, car l'information pertinente pour telle décision est reliée au reste du monde par des liens complexes et multiples, elle n'existe que très rarement sur le mode d'une information pure qui serait déjà délimitée pour le choix. Par ailleurs, le plus souvent c'est l'intention, la visée elle-même qui va donner un relief et un sens particulier à une partie de l'ensemble que constitue le monde. C'est au sujet de délimiter, sous la forme d'un « petit monde » le sous-ensemble des états du monde pertinent pour sa décision. Ici nous sommes assez proches de deux idées de Savage : tout d'abord que le concept de monde doit être taillé à la mesure de la situation décisionnelle, et, en relation directe avec ce qui précède, que toute décision ne peut se situer que dans un microcosme³.

Enfin, on ne saurait analyser le concept d'environnement sans au moins évoquer l'orientation immédiate⁴ que lui donne le concept d'altérité et de société. N/M (1947) dans le

¹ Il faut aussi bien noter que toute l'information n'est pas disponible sur le même mode comme le souligne Akerlof (1991), p 2. L'information « saillante » (« salient ») aura tendance à avoir une influence plus directe sur le SRD, qui devra être d'autant plus vigilant.

² Voir Romano (2010), p 682-684. Comme le remarque cet auteur, en commentant Heidegger, c'est la totalité qui est première pour la perception, le monde m'est donné comme une série d'ensembles d'objets finalisés.

³ Voir Savage (1954), p 82-91 ; Picavet (1996), p 193-196.

⁴ Orientation qui peut être immédiatement morale comme en témoigne la phénoménologie de l'altérité de Levinas : en apparaissant sur le mode de l'expression ou du visage autrui éveille en moi la responsabilité que j'ai envers lui. Voir *Totalité et infini*, p 211-242.

début de leur livre envisagent une économie de type « Robinson Crusoé »¹ où le sujet ne peut et ne doit envisager que la maximisation de certains indices, en laissant de côté le fait que dans le monde réel les sujets peuvent parfois être en compétition. Or, ce type de mono-société n'existe pas et la présence de l'autre dans le monde sur le mode qui est le sien donne à l'environnement une orientation spéciale, qui peut par exemple être celle du regard de l'autre sur ma décision comme le suggérait Allais². C'est d'abord par l'humain que le monde, de façon générale est orienté, et à aucun moment je ne peux ni de dois dans ma visée faire abstraction de l'atmosphère d'humanité du monde³, atmosphère qui pénètre toute la sphère des objets – y compris le corps comme premier objet de l'environnement – sur le mode de la culture⁴, de la finalité, de l'utilité. C'est le monde en tant qu'il est humain et social qui donne au décider sa délimitation première. Il n'y a d'environnement décisionnel que dans la prise en compte de l'altérité comme source de décision possible avec laquelle je dois composer.

3) La résistance et les orientations de l'environnement.

Comme nous venons de le voir, l'environnement décisionnel se déploie et s'oriente dans pas moins de cinq dimensions. Cependant, il est une caractéristique de ces dimensions que nous n'avons pas encore pris en compte et qui est pourtant très importante : toutes sont susceptibles de produire une certaine forme de résistance dans la prise de décision, notamment dans les phases d'édition et d'évaluation du SRD ; *décision que nous définirons – en adéquation avec la TP – par le travail du SRD dans l'élément de la représentation : construction de la visée suivie d'une édition et d'une évaluation dont le terme dernier est l'inscription dans la réalité sur le mode d'une action*⁵. Nous soutenons ici la thèse qu'il n'y a

¹ N/M (1947), p 9-12.

² Mais c'est aussi sur le mode de la coopération et de la décision partagée qu'autrui peut intervenir dans la construction de mon environnement décisionnel.

³ Pour reprendre une image de Merleau-Ponty (1945), p 405.

⁴ Pour un point de vue intéressant dans le cadre de la psychologie cognitive sur les rapports entre culture et décision, voir M. Donald (2009).

⁵ Nous nous tenons aussi assez près de ce qu'on entend classiquement par décision depuis *L'Éthique à Nicomaque* d'Aristote. La TUA peut d'ailleurs être considérée comme une tentative de formaliser le phénomène de délibération (proaïrêsis), qui chez le Philosophe précède la décision. Le tort de la TUA – qu'évite partiellement la TP – est justement de faire de la délibération la totalité de la décision, puisqu'elle ne parvient pas à formaliser l'édition des perspectives et leur évaluation par le biais du système de référence décisionnel. On peut dire que dans la TUA délibération est synonyme de calcul, ce qui n'est pas en soi une mauvaise métaphore, si on en reste à ce registre de langage, c'est d'ailleurs ainsi que l'entend Aristote aussi : calcul des moyens en vue d'une fin. En ce qui concerne le problème de l'action, qui ne sera pas abordé de façon systématique ici, puisque nous nous situons non pas dans le cadre d'une philosophie de l'action mais en amont de l'action : dans le domaine de

pas de décision qui ne fasse fond sur une résistance minimale, résistance qui est susceptible de degrés, et qui s'exprime de différentes façons.

Il est nécessaire de penser la résistance comme composante de tout processus décisionnel et il faut inclure la résistance dans la définition même du concept d'environnement décisionnel, car sinon c'est comme faire l'hypothèse que le monde et la réalité sont totalement transparents au vouloir du sujet, que par exemple, toute l'information est déjà disponible et déjà organisée, que le temps et l'espace ne sont rien pour le décideur, ce qui est assez irréaliste : c'est justement une grande carence de la TUA de N/M que de faire l'économie de ces éléments. Le subjectivisme de Savage représente à cet égard un progrès en émettant l'hypothèse d'une construction subjective du contexte probabiliste.

Dans le domaine de la décision, il faut distinguer deux types de résistances qui se situent à deux moments distincts du processus : dans le décider lui-même et les représentations sur lesquelles il fait fond d'une part et, d'autre part, dans la réalisation de l'action. Nous nous concentrerons surtout sur les résistances liées à l'aspect représentationnel de la décision elle-même. Une attention particulière sera portée aux concepts de temps, de délai et d'attente, puisque nous montrerons dans notre dernière partie qu'il existe un coefficient de pondération temporel ayant quelques analogies de fonctionnement avec le coefficient de pondération élaboré par K/T dans le cadre de la théorie de la perspective. Le fil conducteur que nous allons suivre dans l'analyse du concept de résistance sera basé sur les différentes orientations de l'environnement décisionnel exposées plus haut.

D'un point de vue classique, notamment platonicien et kantien¹, c'est l'irruption du plaisir ou plus généralement des motifs de la sensibilité qui vient faire obstacle ou obscurcir la prise de décision, particulièrement en ce qui concerne la décision morale. Ces éléments sont à verser au registre de l'effet et de l'influence souvent néfaste du corps sur l'âme ou de la résistance essentielle à mon vouloir que représente la matière en tant que telle². Dans la suite de ce travail, nous aimerions montrer que les premiers motifs de résistance sont dus au

la représentation, on peut se reporter aux réflexions séminales de Ricoeur (1950) ; mais surtout Frankfurt (1978), Davidson (1980), Searle (1983), auxquelles Ricoeur (1986) et (1990) donnent une résonnance particulièrement éclairante ; le livre de J. Proust (2005) donne un éclairage intéressant aux sens qui peuvent et doivent être attachés au concept d'action : où commence et où s'arrête l'action ? L'ouvrage propose aussi une reconstruction du concept de « volition » comme clé de voûte des mécanismes à l'œuvre dans la volonté.

¹ Voir *Protagoras* 352D-353C ; *Fondements de la métaphysique des mœurs*, Première section, où il faut exclure de la sphère morale, tout ce qui déborde le pur respect de la loi.

² Elle peut être pensée par exemple sur le mode de l'antitypie leibnizienne : la matière c'est ce qui résiste.

système de référence décisionnel¹ lui-même puisque son rôle essentiel est entre autres de *stabiliser l'existence décisionnelle des sujets autour de certains axes* comme les croyances et les valeurs.

Si nous partons de la première orientation de l'environnement décisionnel que nous avons mis en avant à savoir la dimension subjective immédiate, deux ordres de résistances sont à analyser. Le premier est dû à la constitution et à la façon dont fonctionne le système de référence décisionnel. Comme nous l'avons vu dans notre introduction, il est un sous-système dynamique du SRG². Il actualise à l'occasion des phases d'édition et d'évaluation ce qui définit le sujet comme ce sujet-ci, par exemple les croyances, les connaissances, les valeurs, ce qui le différencie de tous les autres, de façon irréductible et singulière. Trois types de résistances peuvent être attribuées à ce système :

- L'*inertie*³ – la préférence pour le statu quo⁴ –, qui est liée à l'aversion au risque et à la façon dont les habitudes fonctionnent dans le système du décideur.
- L'*akrasie* qui fonctionne à peu près sur le même mode que l'inertie, la différence étant que nous pensons que l'akrasie doit être reliée encore plus fortement à l'aversion au risque, et surtout au fait – mis en avant par K/T (1979) – que les pertes pèsent plus lourd que les gains.
- La *procrastination* qui est elle aussi tributaire de l'inertie et qui consiste à surévaluer l'utilité du statu quo par rapport au déclenchement de la décision.

Comme nous le verrons dans notre troisième partie, *l'akrasie et la procrastination sont toutes les deux des conséquences de l'inertie du SRD*. Ce sont trois formes de résistances internes de ce système : elles sont liées à la difficulté de modifier son fonctionnement dans une direction qui semble incompatible avec l'état actuel des croyances et des valeurs, ou à la

¹ En tant qu'il est le lieu de certaines représentations.

² Nous faisons l'hypothèse que le SRD fonctionne comme en théorie de la perspective : il édite et code les résultats puis les évalue, plus généralement, il code les buts que se fixe le sujet.

³ L'inertie décisionnelle à laquelle nous faisons référence ici se situe en continuité avec la théorie des heuristiques et des biais de K/T par le truchement du concept d'habitude. Nous verrons que le système de référence décisionnel fonctionne à l'aide d'un certain nombre d'heuristiques.

⁴ C'est ce qui ressort de nos analyses de la TP.

difficulté de quitter ou d'entrer dans un nouveau régime d'habitudes, que le sujet se représente comme une perte par rapport à l'état actuel. L'inertie est la conséquence de la tendance qu'ont les sujets à décider la plupart du temps selon les mêmes critères et à partir du même fonds de croyances et de valeurs¹. Comme nous le verrons il y a un aspect positif de l'inertie : elle nous permet d'atteindre une certaine forme de stabilité qui est le mode d'être des décisions les plus quotidiennes et les plus routinières. L'inertie permet au SRD de se stabiliser et de se sédimenter autour de certaines habitudes de vie ou de pensée : elle fait de nous des êtres cohérents².

En intégrant la représentation des différentes orientations environnementales exposées plus haut comme source de résistance ou de stabilité dans les décisions nous entendons donner une première approximation de la complexité du fonctionnement du système de référence décisionnel. Complexité à laquelle les théories mathématiques et psychologiques que nous avons étudiées ne rendent pas justice.

Le deuxième ordre de résistance du domaine subjectif est à relier à la problématique classique mentionnée plus haut : celle de la résistance du corps³ à la décision. L'analyse de ce type de résistance devrait passer par une compréhension des concepts d'effort, de fatigue, de souffrance ou de maladie, en situant le psychisme du sujet dans le prolongement du corps. Mais nous ne pousserons pas plus avant dans cette direction, car cela nous entrainerait trop loin de l'aspect représentationnel de la résistance⁴ que nous privilégions ici.

Cependant, même si nous laissons de côté l'influence proprement physique que peut avoir le corps dans le processus décisionnel, il y a *l'image du corps propre*, sa représentation qui doit aussi être prise en compte. La façon dont je me représente mon mode d'apparition pour les autres peut avoir une importance centrale et peut être source de résistances qui

¹ Un exemple assez typique de ce type de comportement peut être trouvé dans les décisions automatiques ou habituelles non critiques. Comme nous allons le voir, le SRD utilise toujours la stratégie la plus économique.

² Nous pouvons donc dire que le SRD est un système conservateur, mais qui doit quand même posséder une certaine plasticité pour pouvoir suivre entre autres l'évolution du SRG sur le long terme.

³ Voir à ce sujet Goddard (2005). L'article de J.M. Chaumont : « Le corps du concentrationnaire : la honte et le regard », est particulièrement éclairant, certes de façon négative sur les relations directes entre le corps et le décider, l'idée étant la suivante : lorsque je ne suis plus que mon corps souffrant, lorsque le corps sort de façon définitive de son silence idéal et que le besoin devient le maître absolu, alors je perds tout simplement ma capacité à décider. C'est en ce sens là qu'il faut comprendre dans quelle mesure la décision s'inscrit dans une stricte continuité avec le somatique. On retrouve cette idée dans les classiques de Primo Levy (1947), Robert Antelme (1947), David Rousset (1947), Jorge Semprun (1994).

⁴ Puisqu'il s'agit au fond de montrer, dans la lignée de la TP comment la représentation doit être intégrée aux processus décisionnels.

peuvent venir entraver le fonctionnement du système de référence décisionnel¹.

Le deuxième type de résistance que nous allons envisager ici est lié à *l'espace et à sa représentation*. L'espace est un concept important dans le cadre d'une philosophie de la décision pour au moins deux raisons. Tout d'abord l'établissement de la visée par le SRD que nous mentionnions plus haut s'opère dans un espace qui est à la fois réel et métaphorique : l'espace comme lieu des êtres et des objets, le monde que la visée délimite comme environnement décisionnel, et d'autre part, l'espace métaphorisé et intériorisé qui sert par exemple au sujet à se représenter la distance qui le sépare de son but.

La répartition spatiale des objets du quotidien, dans les environnements proches comme le domicile, le lieu de travail, etc., est caractérisée par une manipulabilité et par le fait que le sujet ordonne totalement ou en partie les objets dans l'espace, c'est le sens même d'avoir ceci ou cela « sous la main ». Il y a donc ici une transparence opérationnelle de l'environnement pour la décision² et pour l'action. Plus le sujet va s'éloigner de ce cadre familier et plus l'espace, réel ou représenté va offrir de résistance dans le processus de prise de décision, va devenir une donnée qui devra être traitée comme un problème et non plus comme un appui, comme c'est le cas dans les environnements maîtrisés par le sujet³.

Ce sont les concepts de proximité⁴ et de distance qu'il va falloir analyser comme métaphorisation de la spatialité. Dans le cadre de la prise de décision ces concepts vont intervenir de trois façons différentes : dans la représentation du but, des résultats, des perspectives, dans la représentation des moyens à mettre en œuvre pour parvenir au but, et enfin dans la représentation de l'altérité, si altérité il y a pour la décision donnée.

Dans le premier cas, nous avons affaire à la représentation du but comme s'inscrivant

¹ C'est ce que nous apprennent par exemple les phénomènes du membre fantôme et de la dysmorphophobie. Le premier est longuement analysé par Merleau-Ponty (1945) pour introduire la distinction entre corps sujet et corps objet. Ce phénomène nous indique que le schéma corporel et le corps peuvent être deux choses différentes, et que mon psychisme peut reconstruire ma représentation du corps à tel point que je sente comme toujours présent le membre qui a été amputé, voir Merleau-Ponty (1945), p 104 -109. La dysmorphophobie peut avoir des conséquences bien plus dramatiques. Tout se joue encore une fois au niveau de la représentation du corps et de la direction qu'elle donne à l'ensemble du psychisme et, par définition, au système de référence décisionnel. C'est une pathologie qui peut entre autres causer l'anorexie mentale, et dans certains cas conduire au suicide. Voir à ce sujet, Ferreri/Godefroi/Nuss/Ferreri (2002).

² Il y a une transparence et même pour des décisions qui sortent du quotidien et de l'habituel, le fait de pouvoir décider chez soi, plutôt que dans un environnement inconnu, voire hostile, change beaucoup de choses, et oriente ou allège – si la décision est grave – le travail du système de référence.

³ Il ne faudrait pas conclure de nos propos que les environnements quotidiens sont exempts de résistances ou de problèmes, ce qui n'est pas le cas. Mais si problème il y a il ne vient pas en tout cas de la distance réelle ou métaphorique du sujet à son environnement, il provient de l'irruption de l'inhabituel dans un contexte habituel, et ce faisant il produit un changement de catégorie.

⁴ Pour une analyse magistrale de la proximité, nous renvoyons à Barbaras (2008), p 318-325.

entre l'obtention immédiate et l'impossibilité : tel but est-il atteignable? À quelle distance suis-je de la fin que je poursuis? Ici, il va y avoir résistance si le SRD évalue la distance au but de façon exagérée ou inversement, ce qui peut entraîner le découragement ou la surestimation des capacités¹ qui sont deux types de résistances à verser au registre d'une représentation erronée de la distance métaphorique qui nous sépare de notre but. Cette représentation biaisée peut être due à la façon dont le but lui-même est proposé au sujet, ou construit par ce dernier². D'autre part, l'évaluation que le SRD va faire de la distance de la fin visée aura une influence évidente sur les moyens requis comme pertinents pour atteindre l'objectif : le SRD investira alors trop de ressources ou pas assez. La représentation biaisée de la distance qui sépare le sujet de son but produit une évaluation biaisée des moyens à mettre en œuvre³ pour l'atteindre.

Passons maintenant à un autre type de résistance que nous permet d'aborder de façon assez naturelle la problématique de la spatialité : *autrui* fait partie intégrante du système de référence général sur différents modes, mais c'est aussi au travers des catégories spatiales du proche et du lointain que le sujet se le représente⁴. La métaphore spatiale peut donc nous donner un point de départ pour comprendre dans quelle mesure la représentation de l'autre peut être pour le sujet une résistance à ses décisions. Autrement dit, la spatialité et la résistance de l'altérité se chevauchent.

La métaphore spatiale permet de déterminer comment le sujet est lié aux autres, il y a par exemple les proches, qui sont les membres de sa famille, les amis, ces autres qui – vivants ou morts – ont contribué à constituer le système de référence général du sujet sur le mode de l'éducation, de l'influence, du partage des croyances et des valeurs, ou contribuent à le constituer sur le mode du rapport quotidien, de l'échange journalier. Cette proximité implique un mode de décider spécial : le proche, le parent, l'enfant, l'ami, résistent à ma décision en ce sens que le décideur doit toujours faire attention de ne pas le pas le blesser, de ne pas lui faire outrage.

¹ Voir Camerer/Lovalló (1999), in Kahneman/Tversky (Eds.) (2000), p 414-423.

² Nous retrouvons ici les caractéristiques principales de l'effet de cadrage, qui peut être provoqué par la façon dont le sujet construit la situation, ou par la façon dont la situation lui est présentée, voir K/T (1986). Rappelons, que dans le cadre de nos hypothèses, la visée en le délimitant construit aussi l'environnement décisionnel, et que cet environnement est le lieu où s'inscrit le but et sa possible réalisation.

³ Ici nous sommes dans un schéma assez proche de l'heuristique d'ancrage et d'ajustement que K/T avaient mis en avant dans leur article de 1974

⁴ Autrui est donc doublement spatial : dans le monde, et comme représentation subjective : par exemple les proches.

Le sujet aura alors plus d'égards et le regard possible de cet autre, son jugement – même clément –, pèsera toujours plus lourd sur mes décisions que le regard de n'importe quel inconnu. Pourtant, le regard de cet inconnu lui aussi n'est jamais neutre, et lui aussi peut et doit peser sur mes décisions¹. Même en dehors de l'intimité, de la proximité immédiate, l'altérité intervient dans la plupart de nos choix, comme le remarquaient Husserl, Merleau-Ponty et Levinas : autrui est toujours déjà là, et il a une influence majeure sur le système de référence décisionnel. Influence qui doit être un peu plus détaillée.

Tout d'abord il faut revenir sur le concept de visée dont nous parlions plus haut. La visée c'est ce que le SRD délimite et découpe du monde dans le cadre d'une décision particulière. La partie du monde que le SRD délimite est constituée d'objets divers et d'information, pertinente ou pas, mais elle peut aussi être peuplée de personnes avec lesquelles j'ai ou non des rapports, avec lesquelles je suis proche ou pas. Or ce statut de personne change la représentation que le sujet se fait de cet environnement : il ne peut ni ne doit considérer l'autre comme un simple objet ou une simple source d'information². La présence de l'altérité dans la constitution de l'environnement décisionnel représente une résistance immédiate, nécessaire et normale parce qu'il s'agit d'une personne.

Reprenons l'exemple du sujet qui décide de partir à l'étranger pour une période indéfinie. Outre le fait qu'il doit construire un environnement décisionnel informatif pertinent pour faire son choix³, il doit compter aussi avec les membres de sa famille. Imaginons que le sujet en question ait un grand-père très âgé qui ne peut voyager, auquel il est très attaché. La présence de cette personne dans l'environnement décisionnel va produire une résistance immédiate qui se manifestera sur le mode de l'hésitation⁴. Et plus cette personne sera importante pour lui, plus il va hésiter : nous voyons ici comment la spatialisation métaphorique de la relation à l'autre sur le mode de la proximité conditionne une certaine résistance. Mais, la représentation d'autrui peut intervenir dans la constitution de l'environnement décisionnel sur le mode de la résistance par une autre voie : la hiérarchie et les rapports sociaux. Nous sommes en relation avec les autres sur un mode horizontal :

¹ Pour aller au fond du problème il faudrait analyser les concepts de retenue, de pudeur, et de honte.

² Et même si c'est le cas, même si j'ai à faire à une personne dont la fonction et justement de me renseigner je l'aborderai par le biais de la politesse qui doit être la marque distinctive de mon rapport à tout autre en tant qu'être humain.

³ Plus précisément il s'agit ici pour le SRD d'éditer l'ensemble des alternatives pertinentes.

⁴ Voir Ricoeur (1950), p180-193, pour une magistrale analyse de ce concept.

l'égalité familiale et sociopolitique *postulée* de tous vis-à-vis de tous, mais aussi sur un mode vertical qui est celui qu'impliquent les différentes formes de hiérarchies sociales¹, explicites ou implicites, réelles ou imaginaires². Hiérarchies qui, par les représentations qu'elles occasionnent, peuvent causer toute une série de résistances plus ou moins importantes que nous ne détaillerons pas, puisqu'elles sont trop nombreuses. Nous en analyserons seulement la forme générale.

Décider dans un cadre familial ou social hiérarchisé³ c'est intégrer dans la construction même de la visée l'autre et le regard qu'il peut avoir sur la décision comme composante nécessaire, et comme source possible de résistance ou d'appui. La hiérarchie introduit dans toutes les phases de la décision un moment passif : l'espace d'une évaluation à venir, d'une critique possible, l'irruption dans l'environnement décisionnel du point de vue de l'autre comme déterminant. Encore une fois ce regard peut être représenté comme proche ou comme lointain, mais le sujet ne peut en faire l'économie, et ne peut que difficilement s'y soustraire. C'est ainsi que les choses fonctionnent dans certains milieux familiaux et sociaux, ainsi que dans la plupart des milieux professionnels, ou dans l'administration des états. La hiérarchie donne une orientation spéciale au système de référence décisionnel : le sujet décide, mais un autre a droit de regard sur l'établissement de sa visée, son édition et son évaluation, et a même le droit d'en modifier le cours par le biais du conseil, de l'avis, de la mise en garde ou de la réprimande. Dans le cadre du choix rationnel en contexte hiérarchique, le sujet n'est pas totalement maître de sa décision. Pour le dire autrement, dans un contexte hiérarchique il n'existe pas de décision qui soit totalement privée.

Dans notre analyse des résistances liées aux différentes dimensions de l'environnement décisionnel il resterait deux éléments importants à étudier : d'une part le rôle de la représentation du temps puisqu'elle conditionne assez largement tout le travail du SRD, ainsi

¹ Il faut garder à l'esprit que le concept de hiérarchie ne joue pas le même rôle dans toutes les sociétés. Par exemple, il serait intéressant d'étudier comment la codification et la hiérarchisation extrême des rapports sociaux influencent et conditionnent la constitution du système de référence décisionnel dans un pays comme le Japon ; pays qui, contrairement à l'idée répandue, n'a pas un taux de suicide supérieur à la moyenne. Dans ce contexte les pratiques de la mort volontaire peuvent donner un éclairage singulier sur la manière dont le regard de l'autre ou des autres constitue une donnée beaucoup plus lourde de sens et de conséquences qu'en occident. Voir par exemple Pinguet (1984) : *La mort volontaire au Japon*.

² Par hiérarchie imaginaire nous entendons par exemple la constitution de la famille autour du « pater familias », qui fut longtemps le modèle dominant en occident. Le vocabulaire est très explicite d'ailleurs puisqu'on parle volontiers de « chef de famille », or il s'agit là d'une relation hiérarchique relevant de l'imaginaire social qui peut avoir une influence parfois assez lourde.

³ À notre sens, il existe très peu d'environnements ou d'espaces sociaux qui ne le soient pas : l'être humain hiérarchise son environnement de façon assez spontanée. Tout est une question de degré.

que l'aspect critique ou non critique de certaines décisions, d'autre part il faudrait approfondir le problème de la représentation de l'information, notamment en reliant ce problème au mode d'apparition de l'information elle-même que nous avons évoqué plus haut. Pour des motivations d'ordre philosophique ces deux analyses seront faites seulement dans la suite de notre travail où nous consacrerons au temps une étude longue – du fait de la place centrale qu'il occupe dans toute cognition – et au problème de l'information une analyse qui sera connectée à celle des phases d'édition et d'évaluation.

Les analyses conduites dans cette partie nous ont d'une part servi à établir un lien inédit entre la rationalité et le concept d'heuristique développé dans le cadre de la théorie de Gigerenzer et Al., d'autre part, à construire philosophiquement le concept d'environnement décisionnel comme la première tâche du SRD dans le processus de choix rationnel. La rationalité, dans la lignée de Simon et Gigerenzer, doit être limitée et délimitée du point de vue de l'environnement et du point de vue des capacités des sujets. La réalité n'offre que très rarement un terrain de décision épuré et idéal comme celui que formalise la TUA ou la théorie des jeux, d'où l'importance de construire parallèlement au concept d'environnement le concept de résistance qui permet de rendre compte de la complexité et de la dynamique des représentations à l'œuvre dans l'environnement décisionnel.

D'autre part, comme nous l'avons remarqué, il ne faut pas associer de façon trop immédiate le concept d'inertie avec un défaut qu'il s'agirait pour les sujets de surmonter de façon systématique. Il y a un aspect tout à fait positif et utile de l'inertie qui permet au SRD de se stabiliser autour de certaines valeurs, croyances et habitudes. Il conviendra donc de montrer comment à la fois le système de référence peut se stabiliser autour de certains pôles décisionnels, et comment il peut évoluer sur le long terme afin que la stabilité décisionnelle ne se transforme pas en sclérose décisionnelle.

B – De la théorie de la décision à la philosophie de la volonté.

1) Décision et représentation.

a – **La déconnexion du décidable et du volontaire.**

L'analyse de la notion d'environnement décisionnel nous a permis de nous rendre compte que certains concepts convoqués dans la définition de la prise de décision sont plus complexes et plus élaborés que le laissent entendre la TUA et les théories issues de l'économie ou de la psychologie que nous avons étudiées¹. Ce qui ne veut pas dire que les processus soient différents selon le contexte, mais que dans le domaine de l'économie les décisions peuvent être représentées en termes de biens ou de résultats monétaires, alors que dans le contexte plus général que nous allons analyser beaucoup plus d'éléments entrent en ligne de compte.

Le moment est venu de commencer à aborder sous un angle nouveau deux questions que nous avons évoquées dans le cours de notre travail, qui sont fortement liées :

- 1) *La déconnexion du décidable d'avec le volontaire* dans la TUA et la psychologie d'une part. Question qui n'a jamais été abordée à aucun point de vue et dont nous commençons ici l'étude.
- 2) *La possibilité de construire une philosophie² de la volonté fondée sur le travail du SRD qui est une généralisation des concepts de fonction de valeur et de coefficient de pondération en théorie de la perspective.*

Commençons par nous interroger sur la déconnexion du décidable d'avec le volontaire, qui, si elle est légitime dans le cadre de théories formelles ne doit pas l'être du point de vue de la philosophie, justement parce que dans le concept de volonté est inclus de façon nécessaire celui de représentation qui donne aux décisions humaines toutes leurs densités, leurs

¹ Edwards (1954) et Ellsberg (1961) avaient eu cette intuition mais ne l'avaient pas développé de façon philosophique, c'est-à-dire conceptuelle, d'autre part, tous deux restaient dans le cadre du bayésianisme. Par ailleurs, pour construire notre propre concept avec ses orientations et le concept de résistance idoine il a fallu aussi aller plus loin que Gigerenzer et Al.

² C'est-à-dire une conceptualisation du problème des connexions entre décision et représentation.

caractéristiques et leurs particularités. Il convient donc d'approfondir les raisons qui expliquent pourquoi la TUA et les théories psychologiques que nous avons étudiées, bien que se situant dans le domaine de la décision, peuvent faire l'économie du concept de volonté.

Si nous revenons à la définition d'« homo oeconomicus » que donne Edwards dans son article de 1954¹, nous voyons immédiatement qu'il s'agit là d'une représentation de l'homme tout à fait épurée et simplifiée que l'auteur entend d'ailleurs critiquer. Le but de l'économie est de pouvoir décrire et prédire les décisions humaines en contexte de marché² à l'aide des mathématiques, pour se faire, cette discipline doit se donner une définition de l'homme qui soit la plus simple possible³ et qui soit elle-même fondée sur des éléments formalisables comme l'ordonnancement et la transitivité des préférences⁴. Dans ce contexte-là, on ne voit pas très bien pourquoi le concept de volonté serait pertinent d'autant plus qu'il est tourné vers les valeurs et que l'économie n'a normalement aucune prétention axiologique⁵. Cela explique du même coup pourquoi la notion est absente chez N/M, qui soulignent dès le début de leur ouvrage qu'ils veulent se donner la situation la plus simple possible, sur le modèle de la physique⁶. C'est donc le but de la TUA lui-même, à savoir la formalisation, qui justifie l'élimination du concept de volonté de la sphère de l'analyse.

Si nous prenons la décision ou l'utilité telles que formalisées par la TUA, nous avons à faire à la strate la plus épurée, la plus simplifiée, mais aussi la plus pauvre⁷ du point de vue de la philosophie comme réflexion conceptuelle sur l'humain et sa faculté de décider. Comme nous allons le voir bientôt, ce qui permet de passer d'une théorie de la décision à une

¹ Edwards (1954), p 380-382. On peut aussi se reporter à la définition du comportement économique proposée dans Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2002), p 159. Il s'agit ici d'un article important sur l'aversion au risque, l'effet de dotation et le concept de statu quo qui sont trois notions que nous allons revisiter d'une façon philosophique.

² Le paradoxe de Saint-Pétersbourg et le phénomène de l'utilité marginale décroissante nous montrent que les limites de la prédictivité en contexte de choix rationnel ont été sondées très tôt, voir notre introduction du chapitre I.

³ C'est avec Richard Thaler (1980) que l'économie commencera à intégrer des éléments plus significatifs en ce qui concerne les représentations subjectives, ce sera alors la naissance de l'économie comportementale, qui intègre les résultats de la TP.

⁴ Rappelons que ces deux éléments constituent la définition classique de la rationalité dans le cadre de la TUA. Voir par exemple Ramsey (1926) ; Edwards (1954). Or, avec les analyses de Gigerenzer et Al. nous avons vu que ces deux éléments étaient insuffisants puisqu'ils ne permettent pas de capturer l'aspect écologique de la rationalité, c'est-à-dire l'exploitation de l'environnement par le sujet.

⁵ En tout cas de façon première ou immédiate.

⁶ Voir N/M (1947), p 20-24. Cette idée est aussi présente chez Ramsey.

⁷ Hormis la considération du point de vue formel en tant que tel, qui a un intérêt certain, que nous n'avons pas manqué de souligner dans les parties précédentes. Simon saluait d'ailleurs la TUA comme un des aboutissements théoriques majeurs du XXe siècle. Il faut aussi rappeler que l'aspect formel capture tout de même des éléments importants, si nous réfléchissons à une clause comme la transitivité, elle semble quand même assez proche de l'intuition. Il en va de même pour l'axiome d'Archimède et le PCS dans certains contextes précis comme nous l'avons vu plus haut. « Pauvreté » est ici à entendre dans son sens hégélien, comme condition de possibilité d'enrichissement par le travail du concept.

philosophie de la volonté c'est le concept de représentation, ou pour le dire autrement : *la volonté c'est la décision en tant qu'elle s'inscrit dans l'élément de la représentation*¹.

Nous pourrions dire finalement que la déconnexion du décidable d'avec le volontaire dans la TUA provient avant tout d'une application du principe d'économie² dans le but de la formalisation, ce qui est parfaitement légitime étant donné la difficulté de la tâche qui consiste à axiomatiser de façon aussi précise que possible la décision humaine. Et, dans ce domaine-là, malgré les réserves et les paradoxes nous pouvons dire qu'elle a atteint son objectif principal et qu'elle a ouvert la voie à de nombreuses tentatives de formalisation de l'utilité et de la décision.

Seulement la philosophie ne saurait se satisfaire de l'aspect seulement formel de la décision que propose la TUA : il faut arriver à penser le contenu représentatif à l'œuvre dans le choix rationnel dans toute son extension et sa double corrélation : avec le monde soumis au principe de représentation et avec le tissu cognitif du sujet³. Le principe d'économie et la motivation empirique permettent aussi de comprendre pourquoi le concept de volonté n'émerge pas des critiques de la TUA que nous avons étudiées. La théorie d'Allais et les idées d'Edwards représentent certes un pas vers le réalisme et vers l'introduction des dimensions psychologiques dans la prise de décision, notamment en ce qui concerne l'évaluation des perspectives et la pondération des probabilités⁴. Cependant il faut garder à l'esprit que malgré cette avancée, les expériences qui se sont inscrites dans leur continuité ont été menées pour la plupart dans un environnement *totalelement artificiel*, désincarné, et même pour celles – comme dans le cadre de la TP – qui ont été menées dans le monde réel sur le mode du questionnaire hypothétique, rien ne garantit que les sujets choisissent les mêmes options dans la vie réelle. Il y a donc une *artificialité du protocole expérimental* et du questionnaire hypothétique qui aux yeux de Gigerenzer (1991) par exemple disqualifie bon nombre de résultats, en particulier concernant les biais cognitifs⁵.

¹ C'est ce qui fait toute la différence entre la décision humaine et la décision machinique.

² Dont l'énoncé classique est : « Entia non sunt multiplicanda praeter necessitatem ».

³ L'éventualité de décisions asubjectives ne doit pourtant pas être écartée, notamment dans le domaine de l'informatique théorique et dans celui de la cognition animale.

⁴ Dans chaque cas il y a en fait deux critiques juxtaposées : une critique de la négligence par la TUA de certains éléments psychologiques qui donne lieu à une critique du formalisme lui-même. C'est particulièrement évident dans le cas d'Allais (1952) et (1953) qui propose d'ailleurs un formalisme original.

⁵ À ceci il faut ajouter que comme le remarquait Morgenstern (1979), p 178, le questionnaire qui est à la base du paradoxe d'Allais proposait des alternatives tout à fait irréalistes aussi bien au niveau des rétributions que des probabilités. On peut lire en filigrane dans cette critique que le paradoxe et les incohérences sont provoqués par la façon dont le questionnaire est établi.

Par ailleurs comme les expériences ou les questionnaires furent élaborés pour tester un modèle¹, à savoir la TUA, il n'y avait aucune nécessité – toujours en vertu du principe d'économie – d'introduire un élément qui n'était pas inscrit originellement dans la structure mathématique. Pour le coup, si des éléments nouveaux furent ajoutés au fur et à mesure pour enrichir la théorie, comme en témoigne la TP, c'est parce que celle-ci ne parvenait pas à capturer l'aspect représentationnel de la décision, mais tout s'est joué dans la confrontation des axiomes avec l'expérience, rien n'a été ajouté de façon arbitraire². Et c'est justement par cette confrontation qu'ont pu être mis à jour et formalisés les deux concepts centraux que sont l'édition et l'évaluation, notamment grâce aux phénomènes d'aversion au risque et de pondération des probabilités.

Il y aurait peut-être un dernier aspect à mettre en avant qui tient au concept de volonté lui-même, et qui peut en partie avoir motivé sa disqualification de la TUA et d'une partie de la psychologie³, c'est son lien avec une tradition philosophique et métaphysique très chargée qui se laisse difficilement appréhendé par les mathématiques ou par l'expérience⁴. D'autre part, le fait que la volonté soit liée de façon immédiate avec les valeurs et la liberté alourdit encore la tâche : formaliser la volonté conçue de façon classique, comme la faculté qu'a l'homme de s'autodéterminer d'après des principes, ce serait du même coup essayer de formaliser la liberté et l'axiologie. Enfin, *last but not least*, la philosophie occidentale⁵, non contente d'avoir constitué le concept dans une tradition métaphysique liée à une théologie⁶, propose une vision de la volonté qui est très loin d'être unifiée. Certaines de ces théories ne prennent pas du tout en compte les données scientifiques qui étaient ou qui sont maintenant à notre disposition⁷.

¹ Pas toujours de façon directe, comme nous l'avons vu avec les travaux de Slovic/Lichtenstein (1968) et Tversky/Kahneman (1974).

² Ce dont témoigne par exemple l'introduction du concept d'ambiguïté par Ellsberg (1961), à partir de son expérience de pensée.

³ Un auteur comme Ainslie (2005), p 635, pense que la construction du concept de volonté a pu être vue comme « démodée » au XXe siècle.

⁴ Il est d'ailleurs assez significatif que le seul philosophe qui ait intégralement mathématisé ses intuitions, à savoir Spinoza, élimine de son système le concept de volonté et de libre arbitre. Par ailleurs, La thèse classique depuis *Œdipe*, le chant XII de l'*Odyssée* et le stoïcisme impérial qui veut que l'on ne puisse faire l'expérience de la volonté que dans son exercice ne nous mène pas très loin si nous ne rentrons pas de façon précise dans les mécanismes décisionnels soumis à la représentation qui constituent cet exercice lui-même comme volonté. Nous allons montrer sous peu que l'unité du concept subsume sous lui des éléments et des processus multiples.

⁵ Il en va bien entendu différemment pour les philosophies orientales. Pour un point de vue général voir Farago (2002), p 133-144.

⁶ Un témoignage majeur de cette tendance se trouve dans le *De Trinitate* et le *De libero arbitrio* de Saint Augustin.

⁷ Le XXe siècle offre à cet égard un tableau assez intéressant en ce qui concerne les conceptualisations de la volonté. On peut penser par exemple à Ricoeur (1950) qui propose une véritable eidétique de la volonté fondée en partie sur la phénoménologie de Husserl, cependant la théorie de Ricoeur refuse d'admettre la notion de « fait psychique » comme pertinente. Pour Ricoeur

Notre option théorique est justement de partir des acquis mathématiques et de leurs insuffisances philosophiques pour élaborer nos hypothèses : il s'agit en fait de construire une structure conceptuelle à partir des acquis de la TP, acquis qui n'auraient jamais vu le jour sans l'élaboration de la TUA.

Dans un autre ordre d'idée nous pourrions nous demander pourquoi par exemple ne pas abandonner le concept de volonté pour celui de *motivation* ? En effet, du point de vue de nos expériences quotidiennes, il pourrait sembler que la notion de motivation est plus pertinente et moins spéculative que celle de volonté. Or il n'est rien, et ce pour des raisons liées à la difficile définition philosophique du concept de motivation. Tout d'abord la motivation est tournée vers la réalisation de l'action, alors que pour nous la volonté est tournée vers la représentation de la décision qui conduira ou pas à l'action. La motivation est conçue comme une force qui peut se manifester de façons très différentes et dont la définition générale pourrait être : « To be motivated means *to be moved* to do something »¹. Par ailleurs cet aspect dynamique n'a pu donner lieu pour le moment à aucune théorie qui soit vraiment unifiée, « motivation » est un terme générique qui subsume sous lui une multiplicité de phénomènes difficilement réductibles les uns aux autres². Bien que nous soyons d'accord pour intégrer la motivation dans le processus décisionnel, notamment à son terme comme l'interface entre la décision et l'action, nous pensons que la seule motivation sous-détermine complètement la volonté.

Si nous devons continuer à user du concept de motivation dans notre travail ce sera pour désigner la force métaphorique que produit le travail du SRD ou processus décisionnel³

« ... la notion de fait psychique est un monstre. » (1950), p 12. Cette position pose problème puisqu'elle interdit de prendre en compte une psychologie scientifique et mathématique comme fondement d'une enquête sur le vouloir, or c'est précisément ce que nous nous proposons de faire ici. Par ailleurs les concepts de fonction de valeur et de coefficient de pondération pourraient parfaitement trouver leur place dans une eidétique du vouloir, c'est la piste que nous avons initialement envisagée. D'autres points de vue, plus orientés vers une explication causale de l'action sont ceux de Davidson (1970) à partir du rôle des croyances et des désirs – ces deux éléments n'étant pas limitatifs – ou de Searle (1983) à partir de la construction du concept d'*intention en action*. Ces deux perspectives se situent dans la tradition analytique et ont donné lieu à des débats intéressants comme en témoigne Proust (2005), p 57-107. Signalons aussi l'important recueil de Stroud/Tappolet (Eds.) (2003) qui – partant entre autres des considérations davidsoniennes sur la faiblesse de la volonté – propose une réflexion sur les conceptions les plus modernes de la volonté.

¹ Ryan/Deci (2000), p 54.

² Voir Fenouillet (2009) qui dénombre pas moins de 101 théories de la motivation ! Il y a eu cependant des tentatives fructueuses d'unification dans le contexte de la biologie évolutionniste par Bernard/Mills/Swenson/Walsh (2006).

³ Nous ne récusons pas pour autant le fait qu'il existe des motivations somatiques inconscientes ou à la limite de la conscience, qui sont donc liées ou non au concept de besoin, voir Ryan/Deci (2002) ; et Bernard/Mills/Swenson/Walsh (2006). Mais cela permet justement de mettre en avant l'amplitude sémantique du concept, puisque par exemple il existe la motivation comme rapport au soi et à l'estime de soi, qui est à verser au registre de la représentation, et qui est celle que nous

et qui dispose le sujet à agir ou à ne pas agir, en tout cas qui dispose le sujet à une certaine forme de comportement. Nous ne pourrions accepter dans notre théorie qu'un type de motivation qui peut être relié à l'ensemble du travail du système de référence décisionnel, c'est-à-dire au travail de la représentation. Cette motivation sera à considérer comme le dernier maillon de la chaîne qui va de la visée et de la délimitation de l'environnement au codage des buts et des moyens de les atteindre. Notre idée est donc que la motivation est ce qui permet le passage à l'acte, mais nous aurons l'occasion de nous rendre compte que la motivation peut faillir et permettre ainsi au sujet de différer l'acte voire la prise de décision elle-même momentanément ou indéfiniment comme c'est le cas dans la procrastination¹.

Enfin, si nous devons connecter immédiatement le concept de motivation avec les théories critiques de la TUA que nous avons étudiées jusqu'à maintenant, notamment celles d'Allais et la TP, nous pourrions émettre l'hypothèse que dans la plupart de leurs choix, c'est l'*aversion au risque* qui motive les décisions des sujets. C'est le seul lien immédiat qu'il y ait entre ces théories de la décision et le concept de motivation. Or nous allons voir que le concept d'aversion au risque est avant tout à verser au registre du travail du système de référence décisionnel, donc au registre de la représentation.

b – SRG et représentation.

Le problème de la représentation traverse la totalité de la philosophie occidentale depuis Platon, en passant par Kant, Schopenhauer et Hamelin, jusqu'au débat contemporain entre le réalisme et l'éliminativisme dans le domaine de la philosophie de l'esprit par exemple². Dans le cadre de notre travail nous allons nous donner la structure théorique la plus simple pour penser la façon qu'ont les sujets de se représenter l'ensemble de leur réalité décisionnelle : le modèle de l'assimilation que nous trouvons dans le *De Anima* d'Aristote³ et

admettrons ici ; voir Weiner (2005) ; et il y a aussi en amont le type somatique inconscient, qui pourrait être associé au concept de *drive* proposé par Hull (1943).

¹ Nos analyses de ce phénomène s'inscriront dans le cadre d'une réflexion qui aura pour but de décrire certains aspects du travail du système de référence décisionnel sur le temps et la durée. Nous aborderons du même coup le problème du concept de « deadline ».

² Voir entre autres Stich (1983) ; Fodor (1987) ; Drestke (1988). Pour un point de vue précis sur le débat voir la contribution de J. Dokic, in Engel (Dir.) (2000), p 35-62.

³ Voir, *De Anima*, II, 416b 34 – 418b 4 ; III, 429a 13- 429b 24. Aristote pense par exemple la sensation comme l'assimilation au sens spécifique de la forme sensible sans la matière. Ce modèle est celui sur lequel sont pensés les trois types d'âmes : sensitif, intellectif et rationnel.

qui sous-tend aussi bien sa théorie de la sensibilité que sa noétique. Il s'agit de comprendre comment le sujet assimile – au sens étymologique du terme¹ – l'information qui lui vient de l'extérieur par l'entremise de la perception et qui rentre en ligne de compte dans la décision, de penser la connexion entre le SRG et l'information entrante sur le mode de l'assimilation et de l'idiosyncrasie. Cette option théorique a pour but de pouvoir replacer dans une structure philosophique plus large les concepts de fonction de valeur et de coefficient de pondération mais aussi d'expliquer comment les sujets transforment parfois les composantes décisionnelles qui viennent de l'extérieur.

Nous proposons d'associer le terme de représentation au fait d'avoir la capacité de produire et de stocker, à court ou à long terme, des images mentales plus ou moins nettes, plus ou moins fidèles, à partir de ce que nous ressentons, pensons, percevons ou croyons ; et d'associer à nos états intérieurs – qui sont la toile de fond du SRG – une structure sémantique qui nous permet de les verbaliser pour nous-mêmes ou pour les autres : « je suis en colère », « je crois en Dieu », « j'aime lire Kant », « je suis triste », « je dois faire un effort », etc.

Il faut maintenant revenir sur ce que nous entendons par SRG ou système de référence général. Cet éclaircissement nous permettra du même coup de donner une définition plus précise du système de référence décisionnel (SRD) et de comprendre comment les phénomènes d'aversion aux pertes, d'aversion au risque et d'effet de cadrage doivent être intégrés dans un cadre philosophique élargi, fondé sur les concepts de fonction de valeur et de coefficient de pondération en théorie de la perspective. Lorsque nous parlons du SRG nous entendons deux choses. Tout d'abord nous avons en vu l'ensemble des croyances, des valeurs, des connaissances, mais aussi des souvenirs et toute l'histoire personnelle qui constituent la singularité du sujet et font de l'individu un humain irréductible à tout autre. Ce qui signifie aussi que les décisions subjectives seront tout à fait singulières². Tous les éléments qui forment le système de référence général sont de l'ordre de la représentation. Il faut aussi noter que le SRG est soumis à une dynamique qui provient de deux sources : interne, de par les relations et interconnexions des divers éléments entre eux qui peuvent par exemple produire

¹ ad similes : [rendre] semblable [à soi].

² L'idée que nous aimerions faire émerger de notre travail est qu'il n'existe pas dans la réalité deux décisions qui soient semblables, même dans le cas où les résultats soient identiques les éléments qui influencent le processus décisionnel ne peuvent être réduits, même pour des décisions très formalisées de type administratif par exemple. Cela ne veut pas dire que deux personnes ne peuvent pas prendre la même décision pour les mêmes raisons, mais qu'il y a plus que les raisons qui interviennent et que cette addition est irréductible.

des éléments de deuxième ordre : croyance à propos de croyances, justification de croyance, sentiments éveillés par des souvenirs, méta-connaissances, etc. La source de l'évolution dynamique peut aussi être externe, causée par le rapport constant que le sujet entretient avec le monde extérieur et les informations qu'il contient.

Le SRG représente aussi la façon dont le sujet se perçoit dans le présent en rapport avec son corps propre et aux éléments que nous venons d'énumérer, ainsi que par rapport au flux informatif constant mais d'intensité variable qui lui vient de l'extérieur. Dans ce cas-là il peut être assimilée à l'ici et au maintenant du sujet c'est-à-dire sa conscience, conscience qui peut être associée à des sentiments, des états d'esprit, des passions, et toute autre sorte d'états mentaux. Dans le cadre de nos hypothèses la conscience est équivalente à la présence à soi du sujet. Il est très important de bien comprendre que ces états mentaux peuvent être de durées diverses et avoir des influences diverses sur les décisions¹. Par exemple je peux être en colère momentanément, mais je peux aimer la même personne toute ma vie. Ce sont des éléments qui à leur façon viendront influencer la texture du SRD.

Une objection à notre analyse pourrait être la suivante : pourquoi utiliser le même terme SRG pour désigner deux éléments, deux strates du psychisme distinctes à savoir les éléments qui le constituent comme tel et la conscience ? En fait à y regarder de plus près, ces strates ne sont pas distinctes et la conscience peut être assimilée à un épiphénomène du système de référence général que nous avons mentionné plus haut, elle s'inscrit en continuité avec tout ce système. La distinction n'est que métaphorique, il n'y a aucune rupture. La seule différence est une différence d'actualisation : ma conscience est toujours actuelle, tout au moins lorsque je suis éveillé, alors que les éléments du système de référence ne s'actualisent qu'à l'occasion d'événements extérieurs ou intérieurs singuliers qui modifient alors ma tessiture cognitive et ma façon d'être au monde. Autrement dit, tous les éléments du système de référence général ne sont pas actifs en même temps, nous pourrions dire qu'ils sont en puissance et s'actualisent à l'occasion d'une situation particulière.

Dernier point sur le SRG. Nous pourrions à son propos utiliser l'image de la toile² pour

¹ Voir l'article très important de N. Schwarz (2002) in Gilovitch/Griffin/Kahneman (Eds) (2002), p 534-547. Article qui étudie la façon dont l'état d'esprit du sujet, ses sentiments et ses émotions interfèrent avec la prise de décision. Nous étudierons le rôle essentiel des affects et des émotions sur la prise de décision dans notre prochaine partie.

² La métaphore en question nous conduirait à voir le psychisme comme une toile d'araignée, où chaque point est connecté avec les autres, et où ce qui se passe en un endroit trouve écho partout. Pour le coup l'araignée, en contact avec l'ensemble de l'information potentielle serait la conscience. Dans le psychisme, tout se tient.

métaphoriser la façon dont tous les éléments sont reliés entre eux. Bien qu'il y ait des sous-systèmes distincts, il n'y a pas d'élément qui soit isolé des autres. Par exemple croyances, valeurs, connaissances sont liées et peuvent orienter le SRD de façon directe ou indirecte en provoquant certains sentiments, certaines émotions. Les exemples de ce type d'interconnexions pourraient être indéfiniment multipliés. Nous pensons qu'il faut aller plus loin que le holisme de la croyance, c'est l'ensemble du système de référence général qui forme une totalité supérieure à la somme de ses parties, et c'est justement dans cette surdétermination de la totalité par rapport à la somme des éléments que se situe l'irréductibilité du sujet.

C'est à partir de cette structure que va se construire le SRD subjectif. Il s'articule autour de trois fonctions ou opérations distinctes : l'établissement de la visée, l'édition et l'évaluation. Il y a deux éléments qui doivent être soulignés : *certain*s processus décisionnels peuvent être accompagnés d'évaluations probabilistes, *tous* les processus décisionnels se situent dans le temps qui est le véritable élément dans lequel prennent place tous les choix rationnels. Nous faisons l'hypothèse que l'opération de pondération doit être élargie à la phase d'édition du SRD et notamment au phénomène de catégorisation sur lequel nous allons revenir. D'autre part, à partir des opérations d'édition et d'évaluation, qui sont toutes deux issues de la TP, il est facile de connecter le travail du SRD aux concepts d'aversion aux pertes et d'effets de cadrage qui sont centraux pour nous. Le SRD se situe dans l'élément de la représentation et il produit lui-même un certain type de représentation dans le processus décisionnel. Nous allons maintenant examiner de façon plus approfondie la façon dont ce système fonctionne.

2) SRD et fonction de valeur.

a – **Avoir présent et état présent.**

Les promoteurs de la TUA émettaient l'hypothèse que dans l'évaluation d'une perspective, seule la considération des états finaux était considérée par le sujet. Il faut préciser ici que ce n'est pas le fait de prendre en compte les états finaux qui pose problème, puisque c'est ce qui se passe dans la plupart des décisions, mais de prendre en compte *seulement* l'état final sans le rapporter à rien d'autre qui soit de type subjectif. Cette idée avait été par avance

remise en question par le paradoxe de Saint Pétersbourg¹ et la découverte du phénomène de l'utilité marginale décroissante qui montraient que la considération de l'état final dépend en fait de l'état de fortune présent du sujet. Dans la lignée de ces acquis théoriques et de ceux de Markovitz (1952), la TP propose que les perspectives soient codées comme des gains et des pertes par rapport à un système de référence, ce qui permet de relier la perception d'une alternative à l'avoir présent du décideur.

La construction de la fonction de valeur permet du même coup d'introduire, avec le concept de point de référence, les notions de magnitude et d'intensité² que l'on avait déjà croisées chez Allais, notamment dans sa réflexion et son analyse du concept de seuil psychologique³. La fonction de valeur⁴ en TP est une courbe mathématique dont les caractéristiques doivent être rappelées : elle est définie en fonction de la déviation par rapport à un point de référence, elle est concave pour les gains et convexe pour les pertes, sa pente est plus accentuée pour les pertes que pour les gains⁵.

C'est précisément ces trois éléments qui vont nous permettre de définir de façon encore plus précise les caractéristiques et le fonctionnement du SRD, qui joue dans notre théorie le rôle du point de référence dans l'élaboration du choix rationnel. C'est aussi à partir de cette fonction que nous allons analyser la phase d'évaluation du SRD dans notre prochaine partie. Les deux dernières caractéristiques sont des plus importantes puisqu'elles signifient que le sujet est aversif aux pertes et que les pertes pèsent plus lourd que les gains⁶. Il s'agira pour nous de montrer que *le SRD joue dans notre théorie un rôle assez similaire à celui de la fonction valeur en TP, mais dans un contexte élargi aux alternatives non monétaires, c'est-à-dire élargi à l'ensemble des décisions*⁷.

¹ Voir notre analyse du paradoxe, p 20.

² Concepts centraux si on les ramène par exemple aux croyances ainsi qu'à leurs justifications et aux rôles qu'elles jouent dans le SRD.

³ Voir Allais (1953), p 507. Chez cet auteur le concept de seuil psychologique était pensé à partir de la loi de Weber-Fechner. Elle n'est pas convoquée de façon explicite par K/T dans leur article de 1979, même si l'idée qui sous-tend l'analyse de la fonction de valeur est que notre appareillage psycho-sensoriel est plus adapté pour percevoir des changements ou des différences que pour évaluer des magnitudes absolues. Voir K/T (1979), p 32.

⁴ L'introduction de cette fonction dans la TP a donné lieu à une série d'articles qui se situent aux pages 143-209 de Kahneman/Tversky (Eds.) (2000), articles qui explorent les possibilités théoriques associées à l'aversion aux pertes, à l'effet de dotation (« endowment effect »), et au statu quo, sur lesquels nous reviendrons, effets qui sont reliés directement aux caractéristiques de la fonction de valeur.

⁵ Voir le schéma de la fonction de valeur dans notre analyse de la TP ; voir aussi Tversky/Kahneman (1991), p 144.

⁶ Nous renvoyons là aussi au traitement de ces notions dans le cadre de notre analyse de la TP.

⁷ Il s'agit en fait d'unifier ce que la TP distinguait comme trois opérations différentes en trois fonctions d'un même élément : le SRD. Dans le cadre de nos hypothèses, comme nous l'avons vu plus haut, c'est lui qui code les perspectives, ce qui correspond à la phase d'édition de la TP. Par ailleurs c'est aussi le SRD qui est responsable de l'aversion aux pertes des sujets,

Dans le cadre de la TP, et certaines des autres théories que nous avons explorées, « l'avoir présent » (« current asset »), doit être associé à l'état de fortune du sujet ou à ses biens matériels, ce sont les seuls éléments qui sont pris en compte, et c'est ce qui – de notre point de vue – réduit la portée philosophique des ces théories. Cet état de fait s'explique par le contexte théorique des analyses en question : elles se déroulent ou objectivement et seulement en contexte économique, comme c'est le cas de N/M et d'Allais, ou bien lorsque les problèmes sont abordés par la psychologie, c'est en réponse à une problématique économique, comme dans la TP. Il est donc normal que le point de vue soit celui de l'avoir monétaire ou matériel. Dans notre optique il est nécessaire de généraliser les acquis de la TP à des alternatives qui n'engagent pas seulement des biens matériels¹.

Dans leur article de 1979 K/T avaient aussi envisagé que la fonction de valeur puisse être dynamique, c'est ce qu'ils appellent le décalage de la référence (« shift of reference »)². Ce phénomène survient quand les sujets projettent leurs décisions dans le futur, c'est-à-dire lorsqu'ils anticipent les choix qu'ils auront à faire³, ou quand ils sont confrontés à des niveaux de pertes tels qu'ils ne peuvent adapter leur système de référence à l'état présent. C'est ce qui arrive dans le phénomène de « fuite en avant » chez les joueurs pathologiques⁴. Le système de référence ne code plus la perte possible comme telle, il l'élimine de l'alternative et ne garde que la possibilité du gain.

Il est important dans notre travail de pouvoir penser la dynamique du SRD, mais il faudra élargir cette dynamique par rapport aux positions de K/T. Le premier aspect dynamique du SRD réside dans le fait que ce système n'a pas tout le temps le même contenu, car il dépend en premier lieu de la visée opérée en vue de la décision, mais aussi des composantes qui sont convoquées pour construire ladite décision, nous parlons ici des croyances, des

et enfin de la pondération probabiliste et temporelle. Il faut noter au passage que K/T ne se sont jamais réellement expliqués sur la fonction de codage dans la phase d'édition, c'est ce que nous tentons ici.

¹ KT en avaient évoqué cette possibilité dans leurs conclusions, voir (1979), p 42. Les auteurs indiquent tout de même que la théorie peut être appliquée à des « choices involving other attributes, e.g., quality of life or the number of lives that could be lost or saved as a consequence of a policy decision ». Par ailleurs le dépassement du point de vue économique est clairement assumé dans la conclusion de Tversky/Kahneman (1991), in Kahneman/Tversky (Eds.) (2000), p 157.

² K/T (1979), p 40-41.

³ Ce type de prédiction est un problème en soi et a donné lieu à un certain nombre d'études dont la question centrale est la suivante : les sujets sont-ils capables d'anticiper en contexte décisionnel, sont-ils capables de prédire leurs goûts et leurs préférences futures ? Voir, entre autres, Gilbert/Wilson, in L/S (Eds.) (2006), p 550-563. Le problème de l'anticipation sera analysé dans notre partie sur le temps et dans notre analyse de la phase d'évaluation.

⁴ Voir, Blaszczynski/Nower (2001), p 491-494 ; et surtout Branas-Garza/Georgantzis/Guillen (2007), p 130.

valeurs, des connaissances, etc.¹ Nous aurons aussi l'occasion de nous rendre compte qu'à cette dynamique, que nous pourrions qualifier de « en fonction du but » s'ajoute une seconde dynamique située sur l'axe temporel et qui produit une évolution du système sur le long terme. Évolution qui est parallèle à l'évolution du système de référence général.

En contraste avec cet aspect dynamique, il y a l'aspect statique de la fonction de valeur, que K/T avaient aussi mis en avant dans l'article de 1979, et qui a été conceptualisé comme le « statu quo bias »² par la suite. Dans l'esprit des auteurs il s'agit d'un effet de l'aversion aux pertes, dont l'idée est la suivante : le désavantage à quitter l'avoir actuel pèse plus lourd que les avantages que le sujet pourrait en tirer. Il s'agit ici d'un phénomène très important, sur lequel nous allons revenir puisqu'il nous semble pouvoir être relié à une certaine forme d'akrasie. Cependant dans notre théorie, nous refusons d'appeler ce phénomène un biais, car dans le cadre de certaines décisions, quotidiennes, répétitives, identiques quant à certains éléments, prévisibles, le SRD se stabilise autour de certaines habitudes, de certaines heuristiques³ qui lui permettent tout simplement d'économiser ses ressources. Il existe donc une stabilité décisionnelle, liée au fonctionnement du SRD, qui permet aux sujets d'être cohérents sur le court ou le long terme. Cependant il ne faut pas oublier que cette stabilité peut produire les trois types de résistances que nous avons identifié plus haut et qui sont l'inertie, l'akrasie et la procrastination.

Nous pourrions dire qu'il y a presque un aspect paradoxal du SRD puisqu'il est nécessaire qu'il soit à la fois statique et dynamique, parce que ces deux caractéristiques permettent d'expliquer le comportement décisionnel des sujets dans différentes situations à

¹ Il est important de rappeler que le SRD peut synthétiser les éléments, par exemple dans le cadre de décisions habituelles, automatiques et quasi spontanées, qui ne sont pas soumises à un régime délibératif élaboré, le SRD accède spontanément aux éléments pertinents du SRG pour la situation, voir notre schéma introductif, p 219. Et on ne peut plus dire qu'il s'agisse ici de décision, mais plutôt de *manière d'être* ou de *faire*. Mais le SRD peut aussi actualiser les éléments inclus dans le système général dans le cadre de décisions situées dans un régime délibératif : sélectionner telle connaissance, la justification d'une croyance particulière, l'actualisation ou l'acquisition d'une information pertinente, etc. Il est bien entendu par contre que ce n'est pas le SRD qui génère l'état d'esprit dans lequel se trouve le sujet au moment de la décision et qui peut avoir une influence sur cette dernière. Cet état d'esprit est déterminé par la corrélation de la conscience du sujet avec son corps, avec le flux informatif extérieur ainsi qu'avec les éléments du système de référence général qui sont actualisés à ce moment-là. Le SRD ne peut que très peu intervenir dans ce cadre-là, si ce n'est de façon indirecte, par exemple sur le mode de la satisfaction d'avoir fait le bon choix, ou la déception d'avoir fait le mauvais.

² Voir Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 159-170.

³ Il ne faut pas confondre les décisions habituelles avec la spontanéité dont nous parlions plus haut, ce qui fait ici la différence c'est que même dans le registre de l'habituel peuvent intervenir des éléments délibératifs. Par ailleurs, comme nous allons le voir sous peu, dans notre analyse et en contexte décisionnel, heuristiques et habitudes sont liées lorsque le SRD fonctionne dans le cadre de décisions habituelles ou pseudo habituelles. C'est ce que sous-entendaient les analyses de Tversky/Kahneman (1974) et les travaux de Gigerenzer, les heuristiques sont répétitives et par là même leur usage est habituel.

propos desquelles nous allons nous expliquer. Mais pour ce faire, il faut tout d'abord se dégager du concept d'*avoir présent* pour passer à celui d'*état présent*, c'est-à-dire passer du point de vue monétaire et économique au point de vue cognitif et philosophique. Dans notre travail, l'avoir présent est lié à la dimension strictement monétaire et matérielle, il s'agit de ce que le sujet possède en terme de biens ou d'argent à un moment donné. L'état présent lui est tourné du côté de la subjectivité, il représente l'état du système de référence général et par là même du SRD, c'est-à-dire les éléments cognitifs qui font du sujet ce sujet-ci à un moment donné avec toutes ses caractéristiques, croyances, valeurs, connaissances, désirs, etc. Le passage d'un terme à l'autre s'opère en intégrant au concept de fonction de valeur un certain nombre d'éléments cognitifs et représentationnels.

Un phénomène comme l'*effet de dotation* (« endowment effect »)¹ nous indique que même en contexte économique, du fait de l'intervention de la fonction de valeur, la représentation de l'avoir présent dépasse la considération immédiatement monétaire. En effet qu'est-ce que l'effet de dotation sinon l'ajout par le sujet d'une qualité au bien de type représentationnel : parce que ce vin est le mien, même si je l'ai acheté 20\$ et qu'il coûte maintenant 200\$ je ne le céderai pas pour ce prix². Qu'est-ce qui justifie ce comportement si ce n'est ce que rajoute à l'aspect monétaire, la valeur représentée de la possession en tant que telle ? Valeur qui déforme la perception du prix réel, c'est un phénomène qu'il faut noter, car il est important, il montre – ici de façon directe – comment le SRD code un élément décisionnel³ en fonction de l'état présent du SRG. En l'occurrence ici, ce système, du fait de la possession produit une représentation particulière : cette chose est mienne⁴. À partir de cette représentation, qui sera nécessairement convoquée lorsque le sujet décidera de vendre l'objet, le SRD produira une représentation de second degré – une surévaluation – et codera alors le prix du bien en question au-delà de sa valeur⁵ réelle.

¹ Voir Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (2000) (Eds.), p 159-163. L'effet de dotation est objectivement ramené à la façon dont l'avoir est codé par le sujet, en l'occurrence ici, du fait de l'intervention du concept de possession, il est codé au-dessus de sa valeur réelle.

² Ibid., p 159. C'est l'exemple inaugural utilisé par les auteurs.

³ Dans l'exemple original il s'agit de demander aux sujets si ils acceptent de vendre un bien qui leur appartient au prix du marché. Nous nous situons donc en contexte économique de choix rationnel, voir Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (2000) (Eds.), p 159.

⁴ Comme le remarquent Tversky/Kahneman (1991), in Kahneman/Tversky (Eds.) (2000), p 145 : « ...value appears to change when a good is incorporated into one's endowment ».

⁵ Il y aurait d'importantes recherches à mener sur le codage des biens personnels par le SRD et par exemple sur des phénomènes comme le « sunk cost effect », voir Thaler (1999), in Kahneman/Tversky (Eds.) (2000), p 250-252.

Le passage de l'avoir présent à l'état présent va lui aussi s'opérer au moyen du concept de représentation : théoriquement, nous passons de l'avoir présent à l'état présent au fur et à mesure que viennent s'ajouter des dimensions subjectives représentationnelles dans la formalisation des processus de décision. Nous ne reviendrons pas sur ce propos, il s'agit de l'introduction progressive de la subjectivité dans le choix rationnel¹. L'élément à retenir ici est que même dans le cadre de choix strictement économiques – qui représentent la plupart des exemples que nous avons étudiés – interviennent des éléments représentationnels, comme en témoigne l'effet de dotation².

Pour passer de la fonction de valeur sur les options dans le cadre de la TP au SRD il faut tout d'abord sortir du contexte monétaire et économique, qui représente certes une catégorie épurée du décidable, mais qui n'est pas représentatif de l'ensemble de la sphère décisionnelle. Cette opération est justifiée par le fait que même si la plupart des choix que nous faisons dans nos vies peuvent être codés de façon métaphorique ou réelle en terme de pertes ou de gains, ils se situent dans un contexte bien plus large que le contexte économique, et ils ne peuvent être quantifiés par des valeurs monétaires ou strictement matérielles. C'est une *erreur* que de ramener de façon systématique les concepts de gains et de perte au domaine de l'économie, c'est ce que nous allons voir sous peu.

Ce point en amène un autre : il faut que le fonctionnement du SRD puisse englober les décisions de tous types, avec un nombre d'alternatives indéfini³, des alternatives telles qu'elles se présentent dans la vie réelle, c'est-à-dire pas sur le mode du questionnaire hypothétique ou de la loterie. Le SRD doit donc avoir une portée explicative plus générale que la fonction de valeur pour la raison qu'il est connecté à l'ensemble du SRG. Par ailleurs ce passage est légitimité par K/T eux-mêmes qui envisageaient une généralisation de la TP à des options non monétaires⁴. Le problème c'est que la prise en compte d'éléments multiples et dynamiques provenant de représentations subjectives interdit toute tentative de formalisation. Ce qui

¹ À notre sens, c'est Allais dans son article de (1952) qui a eu l'intuition la plus nette de ce phénomène, surtout quand il analyse l'intervention des facteurs secondaires, voir, p 57-60. Nous pouvons dire en effet que ces facteurs sont secondaires par rapport aux quatre facteurs principaux qu'il expose, et surtout par rapport aux distributions de probabilités et à la possibilité de perdre, mais ils jouent néanmoins un rôle central lorsque nous approfondissons la dimension représentationnelle des décisions.

² Comme en témoignent aussi les études réunies dans le chapitre deux.

³ La Théorie Cumulative de la Perspective de 1992 représente la tentative formelle la plus aboutie dans ce sens-là, voir Wakker (2010).

⁴ Voir K/T (1979), p 42-43.

signifie que pour intégrer tous les éléments qui interviennent de façon réelle dans une décision nous devons sortir du formalisme¹.

Un dernier point reste à évoquer avant de passer au problème de l'aversion aux pertes c'est celui de la caractérisation de l'état présent du SRD, comment cet état doit-il être caractérisé ? Tout d'abord, comme nous l'avons vu plus haut, le SRD dans son mode de fonctionnement est tributaire de l'état du système de référence général, c'est-à-dire de toute l'information qu'il contient de façon actuelle ou virtuelle. Cela nous indique que toutes nos décisions ne sauraient être que contextualisées par notre présent². L'avenir ne peut en effet faire l'objet de décision qu'à partir du présent³ et du contenu présent du système de référence général, ce qui peut parfois conduire à des problèmes du type « planning fallacy »⁴ où le sujet évalue mal à la fois le temps qui lui est imparti pour accomplir sa tâche et la quantité de travail que cette tâche va lui coûter. Nous pourrions dire que du point de vue de la décision, hors certains domaines à déterminer⁵, le futur est impondérable en soi et la structure des « mois futurs » demeure imprévisible ou très floue⁶.

Il faut noter aussi que le SRD ne fonctionne qu'à l'occasion de la prise de décision, le *présent du SRD est totalement équivalent à son fonctionnement*, il y a donc des moments lors desquels le système est inactif, à savoir lorsque nous ne sommes pas en train de décider⁷, ce qui est en fin de compte assez rare. De la même manière que Lichtenstein/Slovic avaient démontré que les préférences sont construites à l'occasion des situations⁸, nous postulons ici que le SRD ne fonctionne qu'au moment de la décision. Le SRD intervient seulement à partir

¹ Ou alors l'alourdir considérablement, au point qu'il perde toute capacité prédictive, c'est ce que soulignent Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000) dans leur conclusion, p 170 : « In many cases there is no obvious way to amend the theory to fit the facts, either because too little is known, or because the changes would greatly increase the complexity of the theory and reduces its predictive field ». À notre sens, c'est à ce moment-là que la philosophie peut et doit prendre le relais.

² Le problème de la représentation du futur sera abordé dans la partie sur la pondération et dans la partie qui sera spécifiquement consacrée au temps.

³ À ce sujet, les magnifiques analyses du livre XI, chapitres XIV à XXI, des *Confessions* de Saint Augustin restent d'actualité : il n'y a que le présent qui s'exprime sur le mode du souvenir en présent du passé et sur le mode de l'anticipation en présent du futur, mais c'est le présent dans sa densité qui détermine tout. Nous faisons ici l'hypothèse que la considération du présent en contexte décisionnel a une portée capitale pour les décideurs.

⁴ Voir Buehler/Griffin/Ross (2002), in Gilovitch/Griffin/Kahneman (Eds.) (2002), p 250-270.

⁵ Il s'agit de domaines où l'information statistique est suffisante pour permettre des prédictions plus ou moins précises, nous pensons ici par exemple à la médecine, à la météorologie, voire au sport.

⁶ Voir Tappolet (2010).

⁷ Même pour des décisions qui impliquent plusieurs niveaux et plusieurs étapes, le SRD n'est mobilisé qu'à l'occasion de la réalisation de ces étapes, il est inactif entre elles. Ce que nous voudrions éviter ici c'est l'idée d'un système qui serait toujours en éveil. De la même façon que le système de référence général n'est pas en tout temps totalement actualisé, le SRD n'est pas non plus tout le temps mobilisé, même si nous prenons un nombre très important de décisions au quotidien.

⁸ Voir L/S (2006), p 1-2.

du moment où il faut établir une visée et enclencher à la suite une édition et une évaluation. Mais qu'en est-il alors des décisions automatiques ou à la limite de la conscience que nous prenons quotidiennement, qui sont par exemple la réponse à un signal, du type le feu est rouge, je m'arrête, mais je pourrais tout aussi bien continuer, il y a donc bien une décision puisqu'il y a au moins deux alternatives. Dans ce type d'action et de comportement, *le conditionnement opératoire créé par l'habitude* fait en sorte que ces situations ne sont plus codées comme des décisions, mais comme des *manières d'être* ou de *faire*.

b – L'aversion aux pertes et l'aversion au risque.

Il existe deux phénomènes que nous avons croisé dans le cours de nos analyses et qui interviennent dans le fonctionnement du SRD : l'aversion au risque et l'aversion aux pertes¹. Le SRD lorsqu'il édite les options les code d'une façon à peu près semblable à celle exposée dans la phase d'édition de la TP², à cette différence près que, comme nous l'avons déjà souligné, nous pensons que ces opérations fonctionnent comme des heuristiques, et que parfois le SRD doit produire une règle de son propre fond pour faire face à la situation. D'autre part le fonctionnement du SRD est lié au degré d'aversion aux pertes ou au risque des sujets. Degré qui peut varier d'un individu à l'autre mais aussi pour le même individu selon le type de décision³.

Le phénomène de l'aversion au risque⁴, qui est important dans le contexte de la TP et de la TCP est central pour toute analyse philosophique de la décision⁵. Sa problématisation en contexte économique remonte au moins à l'article de Friedman/Savage (1948) ; le phénomène

¹ Deux phénomènes qu'il ne faut pas confondre bien qu'ils soient reliés puisque l'aversion au risque concerne l'évaluation globale d'une perspective alors que l'aversion aux pertes concerne l'impact qu'une perspective négative peut avoir sur l'état présent du sujet. Il ne faut pas séparer totalement les deux notions non plus puisque c'est la possibilité de perdre (réelle ou imaginée) qui produit l'aversion au risque. Nous allons approfondir les dimensions philosophiques de ces deux concepts. Nous pourrions dire que l'aversion aux pertes est une sous-catégorie de l'aversion au risque.

² Voir K/T (1979), p 28-30.

³ C'est ce que loupe le concept classique d'aversion aux pertes, y compris celui de la TP.

⁴ Dont la forme générale est la suivante : préférer 50 \$ sûrs à un pari ayant comme alternative, ½ de gagner 100, ½ de ne rien gagner, c'est-à-dire préférer la valeur du pari au pari lui-même. Il ne faut pas oublier que les sujets peuvent aussi être neutres vis-à-vis du risque s'ils sont indifférents entre l'alternative de 50\$ sûrs et le pari ayant la même utilité attendue, ils peuvent aussi aimer le risque si ils préfèrent le pari. D'une façon générale les sujets sont aversifs au risque et aux pertes. Voir Peterson (2009), p 179-183. Voir aussi l'article classique de Pratt (1964), et dans une perspective plus récente on peut lire aussi Kimball (1993) ; Rabin/Thaler (2001) ; ces trois articles se situent dans une perspective strictement économique.

⁵ Voir, la préface de *Choice, Values, and Frames*, p XIII. En fait dans le cadre de ces deux théories, c'est le concept le plus important – avec celui de pondération probabiliste – car c'est un effet direct de l'introduction de la fonction de valeur.

de l'aversion aux pertes est lié à la concavité de la fonction d'utilité¹. Dans le cadre des analyses de Friedman/Savage, la problématique de l'achat d'assurance était déjà reliée à l'aversion aux pertes. Ce qui posait alors problème – et ce sera le cas jusqu'à la TP – c'est la compatibilité de l'achat d'assurance avec l'achat de billets de loteries qui par contre indique que le sujet aime le risque. L'axiomatique N/M est insuffisante pour rendre compte de ces phénomènes, et d'une façon plus générale la TUA échoue à rendre compte de l'aversion des sujets au risque². Comme nous l'avons vu, c'est la TP qui résoudra ce problème notamment grâce à l'introduction de la fonction de valeur et du coefficient de pondération sur les probabilités. Afin de pouvoir intégrer l'aversion au risque et l'aversion aux pertes dans notre théorie nous allons exploiter un texte tiré d'un article de Tversky/Kahneman (1991) dans lequel le rôle de la fonction de valeur est connecté à des éléments d'ordres philosophiques qui vont justement nous permettent de comprendre plus finement comment fonctionne le SRD :

« The value function appropriately reflects three basic facts : organisms habituate to steady states, the marginal response to change is diminishing, and pain is more urgent than pleasure. The asymmetry of pain and pleasure is the ultimate justification of loss aversion in choice. Because of this asymmetry a decision maker who seeks to maximize the experienced utility of outcomes is well advised to assign greater weight to negative than to positive consequences»³.

Cet extrait appelle certaines remarques. Il y a trois éléments qui sont liés avec le travail de la fonction de valeur, et par extension avec celui du SRD : la recherche d'une certaine stationnarité ou stabilité des décisions par rapport à un point de référence⁴, l'établissement d'un lien entre le couple gain/perte et le couple plaisir/peine⁵, enfin le fait qui avait été mis en avant dans la TP à savoir que *les pertes pèsent plus lourds que les gains*. Que signifie l'idée que l'existence⁶ des sujets oscille autour d'une certaine stationnarité? Nous avons vu que le

¹ Comme le note Pratt (1964), p 127 : « The aversion to risk implied by a utility function u seems to be a form of concavity, and one might set out to measure concavity as representing aversion to risk ».

² Voir Rabin/Thaler (2001), p 226.

³ Tversky/Kahneman (1991), in Kahneman/Tversky (eds) (2000), p 157.

⁴ Le concept de point de référence est exploré dans Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 165-166.

⁵ Ces concepts mériteraient de faire l'objet d'analyses approfondies.

⁶ Il est important de remarquer que K/T emploient dans leur texte le concept d'organisme, issu de la biologie, et ce afin de marquer l'universalité du phénomène de stabilité. Pour l'introduction du concept de stabilité adaptative à partir de la théorie des jeux, voir Maynard Smith (1979).

SRD pouvait être la source d'une certaine forme d'inertie¹ quant aux décisions, du fait du déploiement de l'existence autour de certaines habitudes, croyances, valeurs, etc., qui forment le tissu cognitif quotidien et qui conditionnent elles-mêmes des manières d'être et des manières de faire. Cela signifie simplement que les décisions que nous prenons la plupart du temps ne sont pas susceptibles de changer radicalement le cours de nos vies, ou de le bouleverser, mais qu'elles visent une modification ou une amélioration qui reste dans un cadre assez délimité. Ce que nous « gagnons » dans nos décisions quotidiennes, reste associé à cette sphère.

Il faut garder à l'esprit que l'existence trouve sa consistance dans une certaine forme de continuité et de linéarité, une neutralité habituelle qui peut être associée au sentiment et à la conscience d'exister elle-même². Dans le contexte décisionnel, le SRD fonctionne en accord et en continuité avec notre mode d'être habituel, il code les décisions par rapport à l'ensemble du SRG tel qu'il se situe au moment des choix, cela assure la cohérence du sujet avec lui-même sur le court terme³. Ce mode de fonctionnement n'empêche pas que le sujet puisse considérer les mêmes alternatives à des points de vue différents, en fonction de références cognitives ou de sollicitations extérieures qui modifient alors la représentation des alternatives⁴, et cela n'empêche pas non plus que le sujet puisse changer ses préférences dans une certaine limite et selon une certaine amplitude déterminée par l'impact de ses affects, de ses émotions et de son humeur, ainsi que par la forme du SRG lui-même. Comme nous le notions plus haut, l'inaptitude du sujet à changer sa manière de décider – c'est-à-dire de prendre une décision d'ordre deux – peut être conçue comme une résistance majeure pouvant mener à l'inertie pathologique qui peut se manifester sur le mode de la procrastination ou de l'akrasie.

Il y a donc un mode de fonctionnement quotidien du SRD qui s'appuie sur l'ensemble du régime habituel et doxique du sujet, sur ses visées immédiates, et sur les relations journalières qu'il entretient avec le monde et les autres. Dans un régime ou le statu quo –

¹ Pour être vraiment exact, l'inertie ne vient pas que du SRD, elle provient de la structure du SRG, et surtout des éléments de cet ensemble à partir desquels le SRD code les décisions quotidiennes : croyances, valeurs, sentiments, etc.

² Phénomène qui peut trouver un écho dans le sentiment immédiat et normal pris à l'existence en tant que telle, au simple plaisir d'exister.

³ Comme nous le verrons dans notre prochaine partie, cela n'empêche pas que les sujets puissent être incohérents.

⁴ C'est ce que mettent en avant de façon très claire les exemples exposés dans Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 165-166. Le changement de point de vue peut amener à recoder une perte comme un gain et inversement. En fait ce qui varie ici c'est ce qui dans l'alternative est mis en relief, nous avons rencontré cela dans L/S (1971).

pensé ici comme une oscillation du SRD autour de l'axe constitué par le SRG – est considéré comme un état satisfaisant¹, la sortie de ce mode d'être peut être codée comme une perte, c'est ce que nous montrera l'analyse de la procrastination². Un caractère central de l'oscillation du SRD autour de son axe habituel c'est qu'elle autorise de petits changements de points de vue pouvant occasionner eux-mêmes des modifications dans l'ordonnancement des préférences et dans la prise de décision. Ce qui signifie que dans certaines limites, qui seraient à déterminer plus exactement, le SRD a la capacité de prendre des décisions de second ordre, et donc de corriger son propre fonctionnement. Le SRD peut donc fonctionner de façon métacognitive.

Le deuxième point envisagé dans le texte par K/T est que le concept de gain soit associé au concept de plaisir et que le concept de perte soit associé à celui de peine. Si l'association fonctionne bien dans le domaine économique, il convient de la questionner en ce qui concerne des décisions qui dépassent ce contexte et de se demander dans quelle mesure le SRD code-t-il tous les choix comme des pertes ou des gains ?

Nos analyses ont montré que la plupart de nos décisions quotidiennes sont déterminées en fonction de l'état normal du système de référence général, état que nous pourrions associer à une certaine forme de neutralité ou de continuité. Il existe aussi des types de décisions qui peuvent nous arracher du quotidien, nous demander un effort particulier, qui engagent une partie ou la totalité de l'existence³, ce sont notamment les décisions critiques ou exceptionnelles. *Que la décision soit quotidienne ou extraordinaire nous pouvons postuler que le sujet pense pouvoir gagner quelque chose en la prenant*, et ce gain ne doit pas être seulement envisagé en terme matériel ou quantitatif, mais en terme qualitatif : le sujet prend sa décision pour bonifier sa qualité de vie, celle de ses proches, ou autre, bref, pour améliorer une dimension de son existence, ou se maintenir dans un état jugé satisfaisant. Et c'est ici que vont intervenir les concepts de plaisir et peine associés par exemple aux concepts de réussite et d'échec que le SRD va coder comme un gain ou une perte, et ce en fonction de leur intégration au système de référence général sur le mode de la valeur. Un sujet dont la représentation de

¹ Voir aussi, Shane (2002), in Gilovitch/Griffin/Kahneman (Eds.) (2002), p 555-557.

² L'idée de considérer le changement comme une perte ressort avec force des analyses de Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 163 : « individuals have a strong tendency to remain at the statu quo, because the disadvantages of leaving it loom larger than advantages ».

³ Les exemples abondent ce type de décisions : déménager, changer de travail, aider un proche ou un ami alors qu'autre chose était prévu, s'engager dans la résistance armée face à un envahisseur, par exemple dans le maquis en France pendant la guerre de 1939-1945, entrer en religion, etc. Nous pouvons penser aussi à des décisions médicales lourdes.

l'échec est intégrée à la toile cognitive comme un véritable drame codera cette perspective comme la possibilité d'une perte maximale, dans le cas contraire, où l'échec est par exemple conçu comme la possibilité d'un deuxième essai¹ il sera seulement codé comme un contretemps possible, l'irruption d'un délai supplémentaire, ou une difficulté momentanée à surmonter.

Dans les processus décisionnels où le SRD édite et code le but comme une amélioration possible, le codage en question peut parfaitement être biaisé : il produit ainsi la représentation d'une amélioration de l'état général mais cette amélioration s'avère être une illusion. Ici nous pouvons reprendre l'exemple du jeu pathologique, où le coup joué en plus représente pour le joueur la possibilité de « sortir » de la mauvaise passe. Des analyses similaires pourraient être conduites en ce qui concerne les addictions au tabac ou à la drogue. Ce qui dans ce cas est codé comme un gain, c'est-à-dire le plaisir de la prise, est en fait une perte dans la réalité².

Avant de passer au problème de la pondération en contexte décisionnel, nous allons nous interroger sur le thème du risque dans une optique philosophique, afin de conclure notre première analyse sur le lien entre le SRD et l'aversion aux pertes. Il ne saurait être question ici de faire une histoire du concept de risque, puisque ce serait du même coup faire l'histoire des probabilités, par ailleurs ces histoires ont été écrites et très bien écrites³.

De la même façon qu'une certaine forme de flou entoure la plupart de nos décisions, du fait que nous vivons dans un monde complexe évoluant de façon permanente, *nous pouvons dire que tous nos choix sont risqués dans une certaine mesure*, mais il faut immédiatement préciser que le risque qui nous entoure de façon constante est minime quant à la possibilité et aux probabilités. C'est ce qu'avait mis en avant notre analyse de la clause de continuité dans l'axiomatique N/M à laquelle nous renvoyons ici⁴. Le simple fait de décider de traverser la rue

¹ Nous excluons bien entendu ici les décisions uniques, non répétables et définitives comme tenter une opération de la dernière chance, etc.

² Dans les addictions, le SRD code les alternatives en subordonnant le risque – qui peut parfaitement être connu et reconnu – au plaisir, qui est en l'occurrence le but. Le SRD utilise ici l'heuristique d'ancrage et d'ajustement : le but est codé comme un plaisir ce qui biaise tout le processus décisionnel. Voir les très pertinentes remarques d'Akerlof (1991), p 3-5.

³ Voir Hacking (1975) ; Daston (1988) ; Gigerenzer et Al. (1988) ; Bernstein (1996).

⁴ Nous reproduisons l'exemple ici : considérons les options suivantes : $u = 1000\$$, $w = 100\$$, et $v =$ être exécuté immédiatement. Il est bien naturel, avec ces alternatives d'avoir l'ordre de préférences suivant : $u > w > v$ et de se dire que quelle que soit la distribution de probabilités sur la loterie $(ua + (1 - a)v)$ jamais je ne prendrais la plus petite chance de jouer

pour aller saluer un ami peut être considéré comme une prise de risque. Nous pourrions dire que l'axiome de continuité parvient à formaliser le fonctionnement normal, voire quotidien du SRD, car dans notre exemple le SRD ne code pas la perspective négative comme une perspective risquée.

Nous pouvons faire l'hypothèse que c'est parce que la possibilité de l'issue fatale ou dramatique est très faible qu'elle est éliminée dans la phase d'édition. En fait si nous regardons les choses de plus près l'alternative négative est rejetée parce que sa possibilité ou probabilité est minime. L'exemple nous montre aussi que c'est la façon dont les alternatives nous sont présentées qui va déterminer le choix des sujets¹.

Il faut donc se demander à partir de quel moment le SRD code-t-il une décision comme risquée, ou plus généralement qu'est-ce que prendre des risques ? Comme nous le remarquons plus haut, l'aversion au risque est liée à l'aversion aux pertes, et c'est à partir du moment où le risque de perdre devient explicite ou tout au moins sa représentation² que le SRD code la décision comme risquée. Acheter une assurance³ c'est acheter un moyen de neutraliser la représentation de la perte, représentation dont l'impact est proportionnel aux primes que les assurés sont prêts à payer⁴.

ma vie contre 1000\$ ou n'importe quelle somme d'argent, aussi grande soit-elle d'ailleurs. Modifions un peu les données du problème de façon à faire l'expérience suivante : quelqu'un me propose de gagner 100\$ tout de suite ou de prendre ma voiture, rouler 5 kilomètres, et récupérer une enveloppe contenant 1000\$. Mon choix ici serait tout à fait différent, et l'axiome fonctionne parfaitement, aussi bien d'un point de vue normatif que descriptif. La réticence originarie vient *l'incommensurabilité théorique* entre les deux alternatives. Mais après réflexion il devient clair que dans la réalité, mes chances de mourir, en effectuant une action susceptible de maximiser mon bien ou mes avoirs, ne sont jamais nulles.

¹ Face à l'exemple issu de l'axiome d'Archimède nous avons un exemple frappant d'effet de cadrage, voir Kahneman/Tversky (1986).

² Ici, il serait pertinent de commencer une analyse philosophique du concept d'*assurance*, qui révèle beaucoup de choses du rapport que les humains ont envers le risque, et comment ce risque est perçu. Analyse qui devrait commencer avec l'émergence des premières compagnies d'assurances sur les biens marchands transportés par voie maritime au XVIIe et XVIIIe siècle. On trouve un écho du début de cette histoire dans *Robinson Crusoé*. L'achat d'assurances est aussi révélateur du lien qui unit les sujets à leurs biens matériels, il est relié de façon assez claire à l'effet de dotation : c'est parce que ces objets, cette maison, ce tableau m'appartiennent que je suis prêt à payer une prime qui dépasse leurs valeurs, encore une fois la possession produit une surévaluation. Les biens que nous possédons sont codés d'une façon tout à fait singulière par rapport à tout le reste, l'investigation de ce lien aux biens immédiats devrait être analysée de façon extensive. Huber/Kunz (2006) appellent « risk defusing operator » les éléments du contexte décisionnel qui permettent de faire baisser le risque, ce sont par exemple les assurances, les vaccins, etc. Ces stratégies sont des heuristiques que le SRD utilise spontanément en réponse à une situation risquée.

³ Comme le remarquent Johnson/Hershey/Meszaros/Kunreuther (1993), in Kahneman/Tversky (Eds.) (2000), p 224-240, l'achat de primes d'assurances exorbitantes provient de l'action combinée d'une perception biaisée des risques et d'une habile manipulation de l'effet de cadrage par les compagnies. Acheter une assurance c'est en quelque sorte acheter sa tranquillité d'esprit face au risque.

⁴ Pour le coup un simple calcul d'utilité attendue avec une assurance voiture classique, en prenant comme pondération les probabilités d'avoir un accrochage, nous ne parlons pas ici d'accident gravissime impliquant la survie des sujets, permet de se rendre compte immédiatement de l'ampleur de l'arnaque... Il en va de même pour les assurances maison, etc.... Le risque est toujours surévalué par les compagnies.

D'un point de vue général, il existe deux façons pour le SRD de coder le risque, deux façons qui peuvent se combiner : selon la possibilité ou la probabilité, et selon ce qui est en jeu dans la décision. C'est pour cela que nous disons qu'il existe des degrés dans le risque, ces degrés peuvent d'ailleurs être reconnus de façon objective, c'est ce que laisse entendre le concept de « prime de risque »¹. À partir d'un degré qui n'est jamais le degré zéro puisqu'il n'existe pas, les décisions peuvent être hiérarchisées à ce double point de vue : jusqu'à la certitude de perdre, jusqu'à la perte ultime, c'est-à-dire la vie ou celle de ses proches qui peut être envisagée comme pire encore². Il serait intéressant, voire essentiel de s'interroger sur l'impact de la représentation de la mort sur le SRD³, de se demander par exemple si notre finitude existentielle est toujours présente comme à l'arrière-plan de nos décisions ? Toutes nos décisions se situent-elles par rapport à cet horizon que constitue l'être-pour-la-mort⁴ ? C'est un élément très difficile à déterminer, car il repose sur les caractéristiques du système de référence général et sur la façon dont le sujet se représente à la fois la valeur de son existence⁵ et le poids de sa finitude⁶.

En effet, il faut toujours prendre en considération que même si il existe des risques objectifs, ou jugés comme tels, qui justement comme nous le notions plus haut peuvent faire l'objet de prime ou de reconnaissance objective, il ne faut pas oublier que le risque dépend surtout, voire absolument de la façon dont le sujet se le représente, et dont il est intégré au système de référence général avec les autres valeurs, ou les autres croyances. La représentation du risque est toujours singulière, c'est d'ailleurs ce qu'indique de façon très

¹ C'est ce qui se passe dans le cas de certains métiers dont la pénibilité ou la dangerosité est reconnue.

² De façon implicite, ce qui est convoqué ici lorsque nous parlons de pertes c'est l'asymétrie des gains avec les pertes, associée au sentiment de plaisir et de peine, nous traiterons cette asymétrie à fond lorsque nous étudierons de plus près le fonctionnement du SRD à partir de l'ensemble des valeurs et des croyances, car pour comprendre cette asymétrie il faut savoir exactement sur quels éléments elle repose, et ce ne sont pas – tant s'en faut – seulement des éléments matériels ou monétaires comme c'est le cas dans la TP.

³ Questionnement notamment engagé par Ricoeur (1950), p 570-578. Il ne fait aucun doute que cette représentation ne soit intégrée au SRG, par exemple sur le mode d'un certain type de croyances. Dans la veine des analyses de Ricoeur, nous pourrions dire aussi que le concept d'effort est aussi le signe de notre finitude.

⁴ Voir Heidegger (1927), p 246-267.

⁵ Dans cette optique il conviendrait d'analyser les concepts de sacrifice ou de suicide rationnel, qui peuvent inverser les perspectives sur le risque : enfermé dans les locaux de la Gestapo à Paris, le 22 mars 1944, après une première séance de torture Pierre Brossolette se suicide pour ne pas risquer de parler. Des cas comme celui-ci nous offrent une matière à réflexion qui malheureusement ne peut être qu'évoquée ici, mais qui devrait faire l'objet d'analyses précises placées sous la question lancinante et introspective : comment peut-on décider de se sacrifier ? La question du suicide rationnel a été abordée – et pratiquée – par les stoïciens, on en trouve aussi une analyse très fine dans le *Hagakurê*, le livre des samouraïs. Toute la problématique de la connexion entre le sacrifice et la décision reste à construire à partir des données contemporaines, c'est d'ailleurs un enjeu en soi, et une reconnaissance de dette envers ceux qui par leur mort volontaire nous ont permis de continuer à penser.

⁶ Ici, c'est un concept tel que la foi qu'il pourrait être intéressant d'analyser et son impact sur le SRD.

claire la classification usuelle en théorie de la décision : les sujets peuvent être aversifs au risque, neutre vis-à-vis du risque, ou aimer le risque. Ces trois comportements – auxquels il faudrait ajouter ceux décrits par K/T (1979) et Tversky/Wakker (1995) basé sur le travail du coefficient de pondération¹ – nous indiquent que la perception du risque est une *idiosyncrasie* opérée entre le risque réel et sa représentation subjective qui peut être on ne peut plus variable. Et cette variabilité est un élément dont il faut tenir compte, même si la plupart des expériences montrent que les sujets sont aversifs au risque en domaine économique², cette perspective ne peut être maintenue dans des domaines plus larges : le rapport du sujet au risque en contexte élargi est fondé sur les représentations intégrées au SRG sur le mode des croyances et des valeurs, qui produisent à partir de leur dynamique propres la réelle, parce que singulière, évaluation du risque : celle qui entrera en ligne de compte dans le codage du SRD.

3) La généralisation du phénomène de pondération : catégorisation et évaluation probabiliste.

a – **Le travail du SRD.**

Le cheminement que nous avons suivi dans les deux premiers chapitres nous a permis de nous rendre compte que très tôt étaient apparus des problèmes liés à la pondération des probabilités en contexte de choix rationnel. L'idée se trouvait déjà dans l'article de Mosteller/Nogee (1951), première tentative qui avait pour objectif de tester la TUA de façon empirique³, elle intervient aussi à différents degrés dans la plupart des articles que nous avons étudiés d'Allais à la Prospect Theory. Cette dernière représente l'aboutissement de ces recherches avec l'introduction du coefficient de pondération qui permet de formaliser mathématiquement comment les sujets se *représentent* les probabilités et calculent l'utilité attendue d'une perspective non pas au moyen de probabilités objectives, mais au moyen de probabilités pondérées⁴. Par ailleurs, comme le montre notre schéma introductif il existe aussi une autre forme de pondération produite par le SRD : la *catégorisation* qui vise à déterminer la situation décisionnelle à laquelle le sujet a affaire.

¹ Voir K/T (1979) ; Tversky/Wakker (1995).

² Voir les articles réunis dans le troisième chapitre de K/T (Eds.) (2000), p 143-209.

³ Voir notre partie sur Edwards.

⁴ Voir Kahneman/Tversky (1979), p 34-38.

La pondération des probabilités doit être reliée à la problématique générale de la représentation, que nous avons commencé à construire dans les parties précédentes et que nous allons continuer à élaborer. D'une façon générale, le phénomène de pondération peut être pensé de deux façons : comme une mise en relief d'un élément dans un ensemble donné¹, comme une distorsion, une transformation de l'information qui s'opère à partir du filtrage subjectif que nous allons détailler dans un moment.

Kahneman et Tversky ont très tôt eu conscience des limites de la théorie de la perspective², et notamment du fait que le coefficient de pondération ne parvenait pas à capturer totalement l'asymétrie du comportement des sujets face aux perspectives positives et négatives. Ce phénomène fut – avec l'intégration de la dominance stochastique et la généralisation à un nombre quelconque de résultats³ – à l'origine de la Théorie Cumulative de la Perspective (TCP) de 1992 dont une des caractéristiques essentielles pour notre propos est le dédoublement du coefficient de pondération en fonction de la perspective d'un gain ou de la perspective d'une perte⁴ afin de rendre compte du fait que les sujets surévaluent les pertes par rapport aux gains et que cette surévaluation a un impact sur les représentations « probabilistes »⁵. Ce qui nous intéresse ici c'est ce que les résultats de la TCP nous laissent entendre, à savoir que la pondération est un phénomène plus général que ce qui avait été originellement établi dans la TP, TP, qui par ailleurs avait montré que la représentation distordue des probabilités n'est pas homogène comme en témoigne le graphe de la fonction⁶. C'est-à-dire que le coefficient n'a pas partout la même valeur : il varie autour de l'axe que représentent les probabilités. Dans le cadre de notre travail, qui est basé sur le lien entre décision et représentation, le phénomène de pondération va être amené à jouer un rôle central et beaucoup plus général que dans le cadre de la TP ou de la TCP, notamment du fait de l'opération de catégorisation lors de la phase d'édition.

Dans notre théorie les questions vont être les suivantes : comment exprimer le phénomène de pondération dans un contexte décisionnel élargi à des choix qui ne sont pas

¹ La pondération est alors reliée au concept d'attention.

² Voir la préface de *Choices, Values, and Frames*, p X-XI, et Tversky/Kahneman (1992), in K/T (2000) (Eds.), p 44-46.

³ Auxquels il faut ajouter le remplacement des probabilités par des capacités de Choquet, et la généralisation au contexte de risque.

⁴ Voir Tversky/Kahneman (1992), p 48, p 50, et p 60.

⁵ Nous mettons le terme entre guillemets, car il s'agit de capacités dans la TCP.

⁶ Voir Kahneman/Tversky (1979), p 37 ; Tversky/Kahneman (1992) ; p 69-75, et Tversky/Wakker (1995).

formalisables en termes de probabilités et de valeurs monétaires ? Sur quels éléments décisionnels porte la pondération ? Comment s'opère le phénomène de pondération et à partir de quoi ? Il s'agira de montrer que la pondération intervient simultanément avec le principe de représentation, autrement dit – et c'est ce qui était en germe dans les analyses du deuxième chapitre – qu'en contexte décisionnel *toute représentation est une forme de pondération liée aux caractéristiques éminemment singulières du SRG*. Par ailleurs, nous allons aussi mettre en avant le fait qu'il existe deux formes de pondération produites dans la phase d'édition du SRD.

Le but premier de notre analyse est de pouvoir ramener la pondération au travail du SRD, et d'unifier par là trois fonctions que la TP séparait à savoir l'édition, l'évaluation et la pondération. Nous proposons ici l'hypothèse qu'il s'agit en fait de trois opérations d'un même module : le SRD. En effet, la TP parvient à expliquer ce qui est pondéré, à savoir les probabilités¹, mais elle ne rend pas compte de ce qui dans le sujet produit la pondération. K/T parlent de la façon dont les sujets se représentent les probabilités, mais n'expliquent pas qu'est-ce qui dans les sujets opère la pondération. Or il faut aller plus loin et analyser ce qui dans le jeu du SRG et du SRD produit la pondération.

Comme nous l'avons remarqué à partir du début de ce chapitre, toutes les alternatives ne sont pas construites avec une mesure probabiliste précise, les perspectives qui se présentent sous forme de loteries sont assez rares dans l'existence. Ce fait nous indique que si nous devons postuler une pondération de la situation décisionnelle par le SRD, elle sera d'une façon générale d'un autre type qu'une pondération probabiliste. Nous voudrions nous arrêter ici sur l'opération de *catégorisation* que fait le SRD, car nous pensons que c'est une forme de pondération, que n'avaient pas vu K/T, bien qu'elle soit liée à la phase d'édition. Elle permet notamment de comprendre deux opérations subséquentes à elle : la contextualisation qui mesure quel rôle la mémoire peut jouer pour la décision présente (contexte connu/contexte inconnu), et la stratégie décisionnelle qui sera mise en place lors de la phase d'évaluation.

L'opération de catégorisation consiste à pondérer la situation décisionnelle selon le type de contrainte qui pèse sur elle, ce peut être par exemple le délai auquel la décision va être soumise, ou l'enjeu de la décision elle-même, mais aussi tout un ensemble de contraintes liées à l'environnement que le sujet de maîtrise pas. La catégorisation est la première opération de

¹ Seulement, comme nous le remarquons dans nos analyses précédentes il apparaît que la fonction de valeur peut elle aussi être considérée comme une forme de pondération portant sur les résultats.

la phase d'édition, elle précède et conditionne la construction des alternatives ainsi que leur évaluation. D'autre part il faut remarquer que c'est dans cette opération que sera évalué le *risque* impliqué par la décision, il s'agira ici d'envisager un autre type de risque que celui exprimé par les mesures probabilistes. Les catégories sont en fait une évaluation subjective de la *situation* dans laquelle se trouve le sujet, pour une décision donnée, il est bien évident que cette situation doit aussi être rapportée à l'état du monde, et que c'est la synthèse de ces deux éléments – intérieurs et extérieurs – qui définit la catégorie dans laquelle se trouve la situation. Nous avons choisi de diviser les catégories décisionnelles en cinq types qui nous semblent assez représentatives des situations qui peuvent se présenter aux sujets dans la réalité, et qui sont les suivantes¹ :

- 1) *Situation habituelle non critique*. Elle appelle peu de commentaires, c'est la situation dans laquelle évoluent les sujets la plupart du temps, dans leur décider quotidien. Cette catégorie amène le SRD à utiliser des automatismes ou des stratégies habituelles qui peuvent se situer aux limites de la conscience. Mais il est important de remarquer que c'est dans cette catégorie-là que travaille le SRD usuellement et qu'il peut trouver une certaine forme de stabilité et de cohérence. Cette stabilisation peut d'ailleurs poser des problèmes et produire une inertie pathologique, dont les dérivées sont par exemple la procrastination et l'akrasie.
- 2) *Situation habituelle critique*. Nous sommes à peu près dans le même cas de figure que pour les situations habituelles non critiques, à ceci près qu'intervient une légère contrainte, souvent une contrainte de temps, un oubli, un élément qui perturbe le cours normal du processus décisionnel ou de l'action mais pas au point d'amener le SRD à produire une nouvelle stratégie. Par exemple il peut s'agir d'un obstacle sur chemin pris habituellement qui amène le sujet à modifier sa trajectoire, d'une boutique que le sujet fréquente habituellement et qui est fermée, une date limite écourtée pour rendre un travail, etc. Dans ce contexte-là l'aspect critique est suffisamment faible pour pouvoir être intégré dans le fonctionnement habituel, les changements éventuels que la situation peut amener sont minimes. Une situation

¹ Voir notre schéma, p 219.

peut aussi être habituelle critique quant à son enjeu, comme c'est le cas pour certaines décisions médicales importantes mais qui n'engagent pas l'intégrité physique du sujet, ni son pronostic vital.

- 3) *Situation inhabituelle non critique*. Ici nous changeons de point de vue, et nous nous plaçons dans une situation où le sujet ne maîtrise pas une partie ou la totalité de l'environnement avec lequel il interagit. Le caractère inhabituel d'une situation peut cependant rapidement d'estomper à partir du moment où le SRD aura construit une nouvelle stratégie décisionnelle pour y répondre et qu'il aura évalué que cette stratégie peut être efficace dans des situations similaires¹. Ici nous pourrions prendre l'exemple d'un sujet qui change de localité ou de travail, et doit par là même reconstruire une partie de ses stratégies habituelles pour faire face aux nouvelles situations qui se présentent. Il est un phénomène important à remarquer – sur lequel nous reviendrons plus loin – à savoir que dans une situation inhabituelle la première opération du SRD sera de tenter d'appliquer une stratégie qu'il utilise dans la situation habituelle qu'il considère comme la plus proche de la nouvelle situation². À partir du moment où nous passons dans la catégorie des situations inhabituelles nous assistons de façon nécessaire à une augmentation de l'investissement du SRD quand aux ressources internes et externes à utiliser dans les processus décisionnels.
- 4) *Situation inhabituelle critique*. La non-familiarité avec l'environnement qui est l'essence d'une situation inhabituelle peut d'ailleurs être produite par la situation elle-même, elle peut en quelque sorte « surprendre » le sujet si le changement s'opère de façon brusque et inattendue. Ici nous gardons les caractéristiques du non habituel mais en rajoutant une ou plusieurs contraintes, qui du fait que le SRD ne travaille pas dans un environnement qu'il maîtrise, aggrave en quelque sorte la situation, et oblige ce système à sortir de sa « zone de confort » d'une façon encore plus drastique que dans les situations inhabituelles critiques. Un exemple de ce type de situation peut être un sujet qui assiste à un accident de la circulation avec blessés, ou une décision médicale à prendre en urgence du fait de résultats

¹ Sur ce type d'auto régulation interne nécessaire des processus décisionnels, voir Proust (2005), p 155.

² Cette idée est assez fidèle aux analyses de Gigerenzer (2001), in Gigerenzer/Selten (Eds.) (2001), p 37-50.

d'analyses inquiétants. La situation est inhabituelle du fait qu'elle court-circuite une partie de la maîtrise que le sujet a sur son environnement, par ailleurs elle est critique, car elle demande une réaction rapide. Ce qui va caractériser ce genre de situation c'est un changement brutal de l'environnement, par exemple passage de la santé à la maladie. Pourtant, dans ce cadre-là aussi le but du SRD sera de ramener la situation inhabituelle à une situation habituelle, dans les limites du possible bien entendu.

- 5) *Situation exceptionnelle*. Comme son nom l'indique, elle ne correspond à rien de connu par le sujet. Si dans le cadre de décisions inhabituelles, le but du SRD est – au moyen de la production d'une nouvelle stratégie – de normaliser la situation, dans le cadre de situations exceptionnelles, le but du SRD est de pouvoir fonctionner sans aucun référent ni situationnel ni décisionnel, sans mauvais jeu de mots nous pourrions dire qu'il s'agit du « mode survie » du SRD. La différence fondamentale de ce type de situation avec les situations inhabituelles c'est que le SRD ne peut à aucun moment envisager de normaliser la situation, il est hors de question de rabattre ici l'inhabituel sur de l'habituel. Ce sont toutes les situations possibles dans lesquelles l'urgence et les enjeux sont maximums, dans un environnement incontrôlable, voire hostile. Dans ce type de situations c'est souvent la vie du sujet ou de ses proches qui est en jeu, ce sont des situations extrêmement graves comme des prises d'otages, des braquages, des situations de conflits armés et de guerre. Comme le SRD est un système adaptatif, et même si ces situations sont rarissimes, il possède en lui les ressources pour répondre – au moins de façon minimale – à ce genre de situations.

Ce qu'il faut retenir de cette typologie, qui est aussi une typologie des décisions, c'est que si nous entendons faire de la catégorisation une forme de la pondération c'est parce que la situation est tout d'abord évaluée en fonction du SRG, c'est-à-dire que cette évaluation est absolument singulière. En effet une décision inhabituelle critique pour un sujet peut très bien être vécue comme simplement inhabituelle pour un autre. Par exemple dans le cas de l'accident avec blessé, si le sujet qui assiste à l'accident est médecin, la situation ne sera pas

catégorisée de la même façon, puisqu'il est pour lui possible de donner une réponse pertinente à la situation d'urgence qui détermine l'aspect critique de la situation.

Par ailleurs il faut noter que plus le sujet va devoir sortir de sa zone de confort décisionnelle et plus il risque d'avoir à faire face en plus de la situation à du stress et à tout un ensemble d'émotions négatives, et que ces émotions vont être amenées à jouer un rôle de premier plan, à la fois dans la phase d'édition et dans la phase d'évaluation. Les catégories déterminent aussi dans quelle mesure le sujet peut compter sur sa mémoire et sur ses décisions passées pour construire une stratégie pertinente¹. Dans le cas où le sujet peut faire appel de façon importante à sa mémoire et à ses ressources nous dirons qu'il édite la décision en contexte connu, sinon il édite sa décision en contexte inconnu et sa mémoire ne peut que très peu, voire pas du tout intervenir.

Enfin, plus nous allons vers de décisions inhabituelles et critiques et plus le sujet aura tendance à tenter d'obtenir de l'extérieur ou de construire lui-même une mesure probabiliste qui soit fiable. Et c'est là un phénomène important qui vient nuancer les vues de la théorie classique de la décision. En effet cette dernière postule que tous les types de décisions arrivent avec des mesures probabilistes, or nous soutenons ici que ce qui rend nécessaire l'intervention d'une mesure probabiliste précise, c'est tout simplement la gravité de la situation, mais que dans les situations normales de simples estimations vagues suffisent généralement. Ce fait explique aussi les réponses souvent décevantes du point de vue normatif que les sujets faisaient dans le cadre des expériences de K/T sur les heuristiques et les biais. Deux raisons peuvent expliquer ce phénomène que Tversky/Kahneman (1974) voyaient comme une défaillance : les enjeux des questions sont tout à fait bénins, le contexte dans lequel sont faites les expériences, bien qu'inhabituel, n'implique aucun investissement particulier, aucun enjeu grave. Dans ce cadre-là, il est normal que les sujets se soient contentés de donner des réponses automatiques approximatives.

b – Mesure probabiliste et évaluations probabilistes.

Nous nous sommes rendu compte en élargissant notre point de vue sur les décisions humaines que souvent des distributions de probabilités comme celles utilisées par la TUA

¹ Voir notre schéma, p 219.

n'étaient pas disponibles dans la réalité. Appliquer des probabilités objectives à des issues sur le mode de la loterie¹ est une situation qui se présente assez rarement. Allais avait d'ailleurs éveillé notre attention à ce propos en incluant dans ses quatre principes fondamentaux de la décision le fait de la déformation subjective des probabilités objectives, pondération qu'il avait intégrée à son formalisme. D'autre part, l'analyse de l'article d'Ellsberg en remettant en cause la distinction traditionnelle entre les différents types de contextes avait du même coup porté un coup sensible au phénomène, admis par la TUA, de l'utilisation des probabilités par les sujets en contexte décisionnel². La question qu'il convient de se poser c'est : qu'est-ce que les probabilités sont censées saisir de nos décisions quotidiennes ? À notre sens, elles ont pour fonction de capturer l'idée de « chance de³ » qui elle par contre est présente dans la plupart de nos décisions : « avoir une chance de », « combien est-ce que j'ai de chances de » ? Il semblerait que si la TUA ne parvient qu'imparfaitement à saisir cet élément, la TP elle y arrive mieux grâce au coefficient de pondération qui rapproche les probabilités de la façon dont elles sont perçues et construites réellement. C'est la différence entre une *mesure probabiliste* (TUA) et une *évaluation probabiliste* (TP) qui est ici en jeu.

En effet, si les probabilités objectives ne peuvent faire partie – comme nous l'avons vu dans notre deuxième chapitre – de la prise de décision, simplement du fait que les sujets ne maîtrisent pas l'outil probabiliste ou statistique, le concept de « chance de » qui en est la représentation métaphorique – nécessairement floue – intervient lui par contre de façon régulière. Il intervient parce qu'il correspond au mode sur lequel le SRD anticipe dans la phase d'évaluation, à partir de toute l'information disponible, la possibilité⁴ pour un événement d'advenir, ou pour un but d'être réalisé. Car, par définition, toute décision est toujours orientée vers un avenir proche ou lointain, comme nous le notions plus haut, le travail du SRD s'opère toujours au présent, mais c'est un présent tourné vers le futur.

Si nous revenons au concept de « chance de », nous pourrions dire qu'il s'agit d'une forme de pondération originaire de notre rapport au probable, au possible et même au temps, comme nous allons le voir sous peu. Une représentation qui a les mêmes caractéristiques que le coefficient de pondération en TP, celle par exemple d'être une échelle qui a pour limites 0

¹ Voir N/M (1947), p 17-20 ; Granger (1967), p 173, Saint-Sernin (1973) p 72-75, Picavet (1996), p 159.

² Voir, Ellsberg (1961), p 245.

³ C'est d'ailleurs sous le terme de « théorie des chances », « doctrines des chances », que les probabilités sont nées.

⁴ Il faudra d'ailleurs à cet égard redéfinir le concept de possibilité.

et 1 et qui permet de situer le possible, le certain et l'impossible dans le travail du SRD. De la même façon que ce dernier pondère les informations extérieures en fonction de la situation, il pondère aussi les probabilités objectives en fonctions de la décision à prendre, c'est ce que nous enseigne la TP et les études qui furent menées dans sa lignée.

Par ailleurs, le concept de « chance de » n'est pas toujours convoqué dans la décision avec la même intensité : entre « il y a une chance sur deux qu'il pleuve demain » et « combien ai-je de chances de guérir », il existe toute une hiérarchie à construire dans la façon dont ce concept intervient dans nos choix. En fait ce qui advient c'est que plus la décision est critique, inhabituelle, ou exceptionnelle, plus elle engage le sujet avec ses ressources, et plus le SRD va tenter de construire une évaluation probabiliste qui puisse être la plus exacte possible. Nous rejoignons ici le concept de *résistance selon la distance* que nous élaborions dans notre première partie : à partir du moment où l'enjeu d'une décision est majeur, et à partir du moment où la distance du but à atteindre est codée comme une résistance importante, c'est-à-dire à partir du moment où la décision sort du registre de l'habituel et nécessite un travail continu et soutenu du SRD, avec une intégration maximale d'information, le concept de « chance de », et par là même les probabilités vont être amenées à jouer un rôle important. Schématiquement, plus la décision est critique et plus le sujet va solliciter ou produire des évaluations de type probabilistes fiables.

Il y a un deuxième point à examiner en ce qui concerne les probabilités subjectives et leurs rapports aux degrés de croyance, ce qui les différencie d'ailleurs des probabilités objectives. Dans les probabilités objectives l'information vient de l'extérieur, alors que dans les probabilités subjectives elle vient de l'intérieur et elle est en général reliée aux degrés de croyance¹, à cet égard il faudrait préciser le lien entre degrés de croyance et probabilités², et comme nous le remarquons plus haut, étant donné qu'il est difficile d'isoler une croyance de toutes les autres dans le tissu cognitif, nous pouvons nous demander dans quelle limite une mesure probabiliste peut être associée à une croyance isolée.

¹ C'est la thèse centrale du bayésianisme, voir Oaksford/Chater (2007), et Chater/Oaksford (2008). Voir aussi Engel (1984), (1997), et (2001).

² Lien qui peut être pensé de différentes façons, par exemple chez Ramsey, nous allons des croyances et des différences de valeurs entre les états du monde vers les probabilités subjectives, alors que chez Savage ce sont à partir de mesures probabilistes que sont dérivées les croyances conçues sur le mode des préférences entre les événements. C'est ce que formalise l'axiome 4, si eu égard aux mêmes résultats, je préfère le rapporter à la réalisation d'un événement plutôt qu'à la réalisation d'un autre c'est parce que *je crois* que cet événement est plus probable. Ici c'est l'évaluation probabiliste qui produit la croyance.

Nous soutenons ici que la force d'une croyance est soutenue par l'action et les forces combinées de toutes les autres croyances et de leurs justifications, et qu'en retour l'intensité de la croyance ainsi produite renforce la cohérence du système et plus généralement celle du SRG. Il est donc difficile d'extraire un élément de l'ensemble et de tester son intensité de façon isolée. Mais plus encore, nous soutenons ici que c'est l'ensemble des éléments du système de référence général qui doivent être impliqués lorsque nous tentons de comprendre comment nous pouvons avoir des croyances d'intensité variable. Interviennent dans ce processus, les connaissances, les souvenirs, les valeurs, et toute la toile cognitive du sujet, ce qui implique que quand nous tentons de mesurer l'intensité d'une croyance, nous mesurons en fait non pas une probabilité, mais la façon dont l'ensemble du système est investi dans cette représentation cognitive particulière, et comment l'organisation singulière du système en ce point peut être considérée comme une justification de la croyance. Par organisation singulière nous entendons la façon dont divers éléments (souvenirs, connaissances, valeurs, autres croyances, mais aussi sentiments, émotions, voire passions) peuvent être sollicités de concert pour justifier une croyance particulière.

Dans le cadre de probabilités objectives nous allons donc avoir deux éléments qui entrent en ligne de compte : le fait qu'elles soient représentées comme des « chances de » dans le contexte d'une *évaluation probabiliste*, d'autre part, les probabilités peuvent être utilisées comme des *mesures probabilistes*, lorsqu'elles interviennent dans une décision de façon objective. Dans les deux cas de figure les probabilités subissent une distorsion du fait de l'action d'un coefficient de pondération¹, déformation qui produit les comportements que nous avons étudiés dans notre partie sur la TP : une surévaluation des petites probabilités et sous-estimation des probabilités élevées.

Il faut, de façon symétrique, s'interroger sur ce qu'il en est de la pondération des probabilités subjectives. Ici aussi il y a deux cas de figure à considérer. Dans l'hypothèse où le sujet est capable de produire une mesure qui satisfasse les axiomes de Savage, c'est-à-dire une évaluation que l'on puisse faire coïncider presque totalement avec l'échelle des probabilités normales, cette *mesure* subira les déformations prévues par la TP et la TCP. Seulement ce cas de figure est très peu envisageable, car il s'agit d'une idéalisation, que Savage assume

¹ Voir Tversky/Wakker (1995), p 1260.

parfaitement par ailleurs¹. Ce qui se passe dans la réalité est beaucoup plus simple et doit ici aussi être envisagé sur le mode de la « chance de », exactement de la même façon que pour les probabilités objectives, par ailleurs la distorsion par le coefficient de pondération sera identique. Comme nous l'avons vu plus haut, les différences entre les sortes de pondérations interviennent à un autre niveau. *Il convient donc de souligner la différence entre les probabilités telles qu'elles existent comme mesure, subjective ou objective, et l'évaluation probabiliste, toujours subjective.* La « chance de » c'est la façon dont est représentée la probabilité dans le cadre d'une *évaluation*, évaluation qui n'obéira pas aux axiomes de Kolmogorov, et qui n'aura avec les probabilités qu'un rapport d'analogie, mais l'analogie sera suffisante pour garantir que mesure probabiliste et évaluation probabiliste soient soumises à la même pondération, le tableau suivant résume les résultats :

	Probabilités		Évaluation probabiliste (« chance de »)
	Objectives	Subjectives	Toujours subjective
Pondération	Constante : π		

Les analyses menées ci-dessus avaient pour but d'explicitier la possibilité de passer d'une théorie de la décision de type mathématique et psychologique à une philosophie de la volonté qui se constitue comme telle dans l'élément de la représentation. C'est ce concept, que nous voyons émerger très tôt de façon problématique qui assure la légitimité de cette opération du point de vue de la philosophie. Le SRD est un sous-système du SRG – ou système de référence général – du sujet qui opère dans et par la représentation. À chacune de ses opérations, que nous avons aussi commencé de détailler, interviennent des éléments représentatifs divers. Il n'y a pas de décision sans représentation, cette déconnexion est possible dans la sphère mathématique, mais pas dans la sphère philosophique où décision et représentation sont totalement liées et unifiées sous le concept de volonté.

La volonté étant entendue comme le travail du SRD dans l'élément de la représentation, il convenait aussi, à partir des acquis de la TP, de nous interroger sur son

¹ Voir Savage (1954), p 7.

fonctionnement habituel, c'est-à-dire son état présent – qui est connecté à l'aversion au risque – et sur la façon dont il opère la pondération de l'information, qui comme nous l'avons vu peut être double : probabiliste lorsque des probabilités sont disponibles, et pondération via l'opération de catégorisation dans la phase d'édition. Dans le cadre de notre réflexion il était important de montrer que l'axiome de continuité pouvait parfaitement formaliser le rapport habituel su SRD au risque, ce qui nous permet encore une fois de souligner que certains axiomes de la TUA peuvent garder un rôle descriptif et normatif dans un contexte plus général. Nous avons remarqué dans nos analyses précédentes que le PCS et le concept de « small world » fonctionnent aussi dans le cadre de nos hypothèses.

Cependant il reste encore des analyses très importantes à mener qui vont venir approfondir notre étude du travail du SRD. En effet nos hypothèses seraient incomplètes si nous ne nous tournions pas vers une étude du rôle que joue le temps dans les processus décisionnels, car comme nous le remarquons dans notre analyse du concept de résistance : le temps est l'élément de la décision. Il convient donc d'analyser comment le SRD se rapporte à la temporalité dans l'élaboration des choix, et aussi comment il peut le pondérer. Nous faisons l'hypothèse que parallèlement au coefficient de pondération sur les probabilités il existe un coefficient de pondération temporel.

Notre théorie ne serait pas non plus complète si nous n'abordions pas de façon précise le fonctionnement du SRD dans les phases d'édition et d'évaluation. Comme le fonctionnement usuel du SRD est lié à l'aversion aux pertes il conviendra de s'interroger sur ce qui fait que les pertes pèsent plus lourd que les gains, et sur le lien de ce phénomène avec le concept d'akrasie.

C – Temporalité, édition et évaluation.

1) Pondération et perception du temps en contexte décisionnel.

a – **Le caractère essentiellement dynamique de toute décision.**

Le passage de la théorie de la décision à la philosophie de la volonté, à partir du moment où il s'opère par le moyen du concept de représentation, ne peut en aucun cas faire l'économie du concept de temps¹. En effet, comme nous allons le voir, le temps est à verser au registre des éléments représentationnels fondamentaux qui entrent dans l'élaboration de nos choix, par ailleurs il faut d'emblée noter que l'intervention du temps dans la sphère décisionnelle se fait sur différents modes. D'autre part, la perception du temps elle-même peut être distordue et par là même pondérée par différents facteurs qui sont susceptibles d'interférer avec le travail du SRD. *Le SRD est donc un système dont le fonctionnement ne peut pas ne pas être temporel.*

C'est dans la première opération du SRD, à savoir l'établissement de la visée, qu'intervient le temps avec les concepts de durée, de délai, de moment opportun. Ce qu'évalue le SRD de façon spontanée dans l'établissement de la visée c'est si le temps doit être considéré pour la décision présente comme une ressource ou une contrainte.

Le rapport entre le temps et le choix rationnel est un problème qui a été très tôt investi par la philosophie occidentale comme en témoignent les réflexions d'Aristote dans l'*Éthique à Nicomaque* ainsi que celles qui traversent l'ensemble du stoïcisme de Zénon à Marc-Aurèle² notamment de par leur réflexion sur les rapports du temps et la volonté³ par le truchement du concept de « *kaïros* » sur lequel nous reviendrons, car il est toujours actuel. D'un point de vue introductif nous pouvons dire que *le temps est l'élément de toute décision*. D'une façon plus générale nous pouvons dire que *le temps est l'élément de toute cognition*.

¹ Du point de vue de la philosophie le lien entre le temps et la représentation est central, c'est ce que montre entre autres l'Esthétique Transcendantale kantienne et certaines données de la phénoménologie contemporaine dont témoignent les travaux de Husserl (1893-1917), Heidegger (1927) et (1966) et Romano (1999). La réflexion sur le temps comme concept philosophique traverse elle toute la philosophie du *Timée* à Jankélévitch et Sartre. Mais le temps demeure une notion hautement problématique et polysémique comme en témoigne l'écart conceptuel qui sépare par exemple le temps en relativité restreinte et la conscience du temps. C'est une notion très difficile à définir comme nous l'enseigne entre autres le livre XI des *Confessions* ou les *Pensées* de Blaise Pascal. La psychologie cognitive reconnaît aussi le caractère problématique du temps, voir par exemple Hancock/Block (2012), p 269.

² Voir en particulier Goldschmidt (1969) : *Le système Stoïcien et l'idée de temps* ; voir aussi Long/Sedley (2001), p 315-334.

³ Voir Voelke (1973) : *L'idée de volonté dans le Stoïcisme*.

Les théories que nous avons étudiées dans nos deux premiers chapitres ne prennent pas en compte l'intervention du temps dans les processus décisionnels, ou alors de façon seulement allusive, en situant par exemple la prise de décision en rapport à un futur proche comme c'est le cas dans l'axiomatique N/M¹. Nous pouvons penser que c'est parce que la TUA formalise seulement des options singulières et non répétitives. Or, la plupart de nos décisions sont intriquées les unes avec les autres quant à leurs alternatives et quant à leurs issues, et certaines d'entre elles peuvent être répétées. Le bayésianisme de Savage représente à cet égard un progrès par rapport à la théorie N/M en ce sens que le concept de révision de croyances que sous-tend l'utilisation du théorème de Bayes implique un aspect temporel. Par ailleurs, d'autres éléments temporels sont dispersés dans la présentation de l'axiomatique, mais ne sont pas systématisés². Qu'en est-il alors de la TP ? Comme elle se situe dans la lignée de la théorie N/M, elle est confrontée au même problème³. Cependant elle a des conséquences temporelles qui sont par exemple exposées par Loewenstein/Prelec (1993) : « Applications of prospects theory imply that people should like to spread gains out accross different days and to concentrate losses in a single day »⁴. Ces faits sont à relier au phénomène de l'aversion aux pertes et surtout à la surévaluation des pertes par rapport aux gains sur laquelle nous allons revenir sous peu. De son côté la TCP n'ira pas plus loin que la TP en ce qui concerne le temps.

Malgré cette carence assez nette du point de vue du formalisme, il n'en demeure pas moins que le temps est un élément essentiel de toute prise de décision. En effet par l'intervention des concepts de buts, d'issues ou d'objectifs, qui nous forcent à anticiper un futur plus ou moins proche, mais aussi parce que l'ensemble du SRG est dynamique.

Parce que le temps est un facteur qui ne peut être ignoré, des entreprises de formalisation de la dynamique décisionnelle avaient été proposées très tôt, notamment par Ward Edwards dans ses articles de 1961 et 1962. Laissons à cet auteur le soin de définir de façon générale ce qu'est une décision dynamique : « In real life, decisions occur in sequences

¹ Voir N/M (1947), p 19. Cela ne veut pas dire qu'aucune tentative pour penser la dynamique décisionnelle n'ait été tentée dans ce paradigme, un exemple significatif à ce sujet-là est celui de McClennen (1990), pour un tour d'horizon complet sur les différentes tentatives de formaliser les décisions dynamiques, voir Manzini/Mariotti, in Anand/Pattanaik/Puppe (2009) (Eds.), p 239-271.

² C'est le cas lorsque Savage discute le proverbe : « Look *before* you leap », il y a ici une connotation temporelle évidente.

³ Toutefois il faut nuancer le propos, car dans la partie expérimentale, la TP envisage un jeu à deux étages, et le but du jeu en question est justement de montrer que les sujets n'intègrent pas la strate informative qui *précède* leur décision.

⁴ Loewenstein/Prelec (1993), in Kahneman/Tversky (Eds.) (2000), p 572. Les conséquences temporelles de la TP sont développées dans le chapitre 7 du volume *Choices, Values, and Frames*, ainsi que dans l'article de Daniel Kahneman : « Evaluation by Moments : Past and Future » qui se trouve dans le même volume, p 692-708.

and information available for later decisions is likely to be contingent on the nature and consequences of earlier ones »¹.

Il s'agit ici de mettre l'accent sur la façon dont le processus décisionnel produit lui-même une information susceptible de conditionner la ou les décisions suivantes, c'est ce qui se passe la plupart du temps dans l'existence où il est à la fois difficile d'isoler les décisions de toute la dynamique cognitive, ainsi que d'isoler les décisions les unes des autres, car comme nous l'avons analysé plus haut, dans le psychisme toutes les parties se tiennent et se soutiennent mutuellement. Par ailleurs, il est important de remarquer que – lorsque les décisions sont imbriquées les unes dans les autres – les conséquences de nos choix ont aussi un impact sur la constitution de l'environnement décisionnel pour les décisions qui dépendent de ces choix² en ce sens qu'elles peuvent le modifier au fur et à mesure, comme le souligne Gonzalez (2005) :

« Dynamic decision making (DDM) is characterized by the need to make multiple and interdependent decisions in an environment that changes as a function of decision maker's actions, in response to environmental events, or in both ways »³.

Le temps est donc le milieu universel, il est comme le remarque Kant dans l'*Esthétique transcendantale* « la forme du sens interne »⁴, mais il n'est pas que cela, il est aussi le médium de la réalité elle-même et de son mode d'apparition. Or, il peut y avoir des tensions voire des décalages assez sensibles entre le temps vécu et le temps « objectif »⁵. La perception du temps est conditionnée par différents facteurs qui peuvent faire percevoir au sujet une même durée comme courte ou longue, il est intéressant de remarquer aussi que tous les moments d'une même durée ne sont pas codés de la même façon, c'est ce que met en avant l'article de Kahneman (2000), et ce codage particulier va entraîner une mémorisation particulière. Il est donc important de garder à l'esprit que le temps aussi peut subir une distorsion et une

¹ Edwards (1961), p 44.

² Il est bien entendu que dans le cas de décisions imbriquées les unes dans les autres, c'est l'environnement décisionnel et par là même la visée elle-même qui se modifie au fur et à mesure que les objectifs sont atteints.

³ Gonzalez (2005), p 142.

⁴ Voir *KRV*, Seite 108. Pour des remarques intéressantes sur la physiologie contemporaine du temps, voir Hansen/Trope (2012), p 1.

⁵ Voir à ce propos Eagleman (2008). L'article est consacré aux illusions temporelles prenant place dans des durées très courtes, il est intéressant que l'auteur dans son propos réactive la question de la *réalité* du temps, p 133.

pondération assez sensible, et que cette pondération va interférer avec le travail du SRD, dont nous avons vu qu'il est un système à la fois statique et dynamique.

De la même façon qu'il existe un mode de fonctionnement habituel du SRD, il existe en adéquation avec ce mode une perception « naturelle » du temps pour le sujet, qui est liée à l'orientation générale et quotidienne du système de référence général ainsi qu'à la routine, aux décisions habituelles. Dans le cadre de nos décisions routinières et habituelles nous percevons la durée d'une façon plus courte que dans une situation non routinière¹ ou une situation qui nécessite l'intervention de plus de ressources cognitives ou autres². Cependant, le lien du SRD au temps est plus complexe et doit être approfondi par le biais des concepts d'*anticipation* et de *niveau d'attention temporelle*.

Nous aurons l'occasion de nous rendre compte lorsque nous étudierons le phénomène de la procrastination que *le présent peut-être totalement surévalué par rapport aux conséquences se situant dans un futur proche ou lointain*. D'autre part, nous avons vu plus haut que le SRD était totalement équivalent à son fonctionnement, il n'y a pas de décisions inconscientes, mais il peut exister des décisions à la limite de l'automatisme pour lesquelles la répétitivité et l'habitation ont effacé la genèse comme décision au sens strict, puisqu'en contexte décisionnel comme dans un contexte plus général : l'habitude c'est l'oubli de la genèse³.

La structure même des décisions humaines, de par le fait que les conséquences ne peuvent se situer que dans le futur, oblige le SRD à sortir métaphoriquement de son propre présent pour se projeter de façon virtuelle dans l'avenir : il s'agit là du concept d'anticipation. Nous pourrions dire que le SRD ne fonctionne que dans le présent mais en vue de l'avenir, aussi proche soit-il. Ici émergent donc deux concepts importants pour toute philosophie de la

¹ Voir Avni-Babad/Ritov (2003).

² En fait, les choses sont un peu plus nuancées comme le soutiennent Hansen/Trope (2012), p 9. Ces auteurs ont montré de façon expérimentale qu'il y a une sorte de décalage paradoxal entre la perception du temps en train d'être vécu et sa remémoration. Par exemple, la durée d'une décision comportant un nombre important d'étapes peut être vécue comme courte alors que le souvenir de cette durée et de ses éléments sera remémoré comme plus longue qu'elle n'a été vécue en fait. La perception du temps est aussi liée au *niveau d'attention temporelle* du sujet, par exemple dans un contexte quotidien, notre niveau d'attention naturel est en adéquation avec ce que nous avons affaire habituellement, mais en contexte d'urgence ou de stress notre niveau d'attention temporelle va être nécessairement augmenté. Voir aussi Angrilli/Cherubini/Pavese/Manfredini (1997), p 972.

³ Ainslie (2005), va jusqu'à intégrer les manies (se ronger les ongles, etc.) dans ce type de décisions aux limites de la conscience, p 638, et ce en vue de casser l'adéquation entre désirer et vouloir. Il existe des décisions dont on peut dire qu'elles sont voulues, mais non désirées. Un sujet qui se ronge les ongles nerveusement, ne *désire* pas se ronger les ongles, pourtant il le veut.

décision : *l'anticipation et la prédiction*¹. Il n'est pas aisé d'établir de distinction précise entre les deux notions, nous proposons ici de faire de l'anticipation une projection du SRD quant aux buts, aux issues et aux conséquences², et de la prédiction une projection du sujet dans le futur d'ordre plus général et qui a trait à la réalisation de certains événements et à l'évolution ou au changement des préférences. Or comme le montrent Gilbert/Wilson (2000) il est très difficile pour les sujets de parvenir à prévoir de façon effective et juste quelles seront leurs préférences dans le futur³. Cela peut d'ailleurs causer un biais décisionnel qui est assez semblable à celui d'ancrage et d'ajustement : si je prédis ce que ma préférence sera et que cette prédiction soit fausse, tout le processus décisionnel que je suis susceptible d'engager pour satisfaire cette préférence future sera nécessairement biaisé par la fausse prédiction qui sert de point de départ à la décision. C'est la représentation du but qui est ici biaisée.

Les préférences peuvent se modifier sur le long terme, elles suivent l'évolution du système de référence général, mais elles peuvent aussi se modifier en raison de l'état affectif du sujet. À partir du moment où cet état modifie l'état du système de référence général et la façon dont le sujet intègre le flux informatif, les préférences peuvent momentanément se trouver altérées en fonction des affections⁴. Enfin les préférences peuvent se modifier selon les modalités décrites par Slovic et Lichtenstein (1968) et (1971), c'est-à-dire en fonction de l'attribut que le sujet va considérer comme le plus important selon la tâche demandée⁵. Dernier point, les préférences peuvent aussi changer en fonction du travail du SRD, lorsque ce dernier considère le même élément sous des angles différents, et lorsque de par la pondération qu'il opère de l'information, c'est un attribut plutôt qu'un autre qui sera mis en évidence.

L'anticipation est une opération plus particulière au SRD, elle a pour fonction de parvenir à une représentation réaliste de la façon dont l'inscription de la conséquence, du but, de l'objectif dans la réalité va changer l'état du monde pour le sujet et son environnement.

¹ Voir, Gilbert/Wilson (2000), in Lichtenstein/Slovic (2006) (Eds.), p 550-563 ; voir aussi Ariely/Zakay (2001), p 200.

² Il est important de remarquer que dans le cas de l'anticipation, la mémoire et l'information ayant trait aux décisions passées jouent un rôle assez important, voir Weber/Johnson (2006), in Lichtenstein/Slovic (2006) (Eds.), p 397-410. Ce qui signifie que souvent l'anticipation aura la forme d'une induction.

³ Voir, Gilbert/Wilson (2000), in Lichtenstein/Slovic (2006) (Eds.), p 555-557. Il est aussi difficile pour les sujets de prédire de quelle façon leurs décisions vont modifier leurs affects, c'est le problème de la prédiction affective (« affective forecasting ») sur laquelle nous reviendrons dans notre analyse de la phase d'évaluation. Pour un point de vue introductif voir Loewenstein/Lerner (2003).

⁴ Il serait intéressant de détailler par une analyse philosophique précise, comment les différents affects de joie, de colère, de tristesse, mais aussi des passions comme l'amour influent de façon nécessaire sur la construction de nos préférences et par là de nos choix rationnels.

⁵ Voir L/S (1971), in Lichtenstein/Slovic (2006) (Eds.), p 52.

L'anticipation doit être associée à la phase d'évaluation du SRD, SRD qui anticipe en l'occurrence l'impact de la prise de décision et de son orientation dans la réalité. C'est ce que d'une façon métaphorique nous entendons par « imaginer les conséquences de ses actes ». Toute décision a pour fonction première de modifier l'état du monde auquel se rapporte le sujet, ce changement peut être tout à fait minime, comme le fait de choisir ce verre dans le placard plutôt que celui-là, ou ultime : choisir de mettre fin à ses jours. Dans les deux cas une anticipation est au principe de la décision, au principe du *geste*¹. D'autre part, le concept d'anticipation serait un appui assez solide qui nous permettrait de passer dans le registre éthique, car les conséquences des décisions, à partir du moment où elles se situent dans une communauté humaine comme nous l'avons vu dans notre partie sur l'environnement décisionnel, peuvent aussi être des conséquences pour les autres, et par là même engager ou impliquer des effets sur les sujets qui m'entourent. Or c'est par l'anticipation des conséquences néfastes que pourraient avoir ses décisions sur autrui que le sujet entre dans la sphère de l'éthique et de la responsabilité².

Dans l'aspect temporel du travail du SRD il est un autre concept important à souligner c'est celui d'*attention temporelle*, qui est la prise en compte du facteur temporel dans le cours du processus décisionnel. Ce niveau d'attention sera bien entendu tributaire de la tâche du SRD et de la complexité des opérations qui sont impliquées. C'est aussi de par le niveau d'attention temporelle que le sujet sera capable de prendre sa décision au bon moment, et c'est ici que nous rejoignons le concept stoïcien de « *kaïros* » qui signifie dans cette école le moment où l'action doit s'insérer dans le monde³. Repousser indéfiniment ce moment c'est

¹ Le geste serait d'ailleurs un concept qu'il faudrait analyser pour lui-même, car son mode d'inscription dans le monde comme trace d'une intériorité permet de remonter de son effectuation à la décision dont il est aussi le signe.

² L'anticipation est un concept important dans la sphère éthique, car il est aussi lié à la procrastination. D'une façon plus générale, il serait intéressant de s'interroger sur le concept de procrastination dans la sphère de l'éthique, en particulier sur la procrastination des états face à des enjeux aussi importants que le changement climatique. La question est abordée par Vallérie Voyer dans son mémoire de maîtrise (en cours de rédaction), dans la lignée des travaux de Steel et de Tappolet. Dans ce contexte-là il semblerait que la procrastination consiste à déléguer les enjeux environnementaux aux générations suivantes, et ce depuis un certain temps, ce qui a pour conséquence de repousser l'horizon de l'action de façon indéfinie.

³ Cette idée nous laisse entrevoir que dans la phase d'évaluation du SRD va intervenir la représentation anticipée du moment de l'action, et que ce moment constitue une part de la situation décisionnelle en tant que telle : à quel moment la décision doit-elle s'inscrire dans le monde pour produire l'action et son lot de conséquences ? Ce phénomène vient confirmer l'idée que sous-entend la maxime d'obédience stoïcienne : « un temps pour chaque chose ». Le SRD anticipe en plus du moment de l'action le moment des conséquences de l'action ou de leur étalement dans le temps. Dans ce contexte, un certain nombre d'études ont montré que les sujets faisaient montre d'une certaine *impatience* : une conséquence obtenue rapidement semble plus intéressante qu'une conséquence meilleure obtenue à plus long terme, voir Ainslie (2005). Cela nous indique qu'il y a comme une tension dans le SRD entre l'anticipation du moment opportun pour l'action et l'anticipation du moment opportun des conséquences de l'action. Le fait de préférer une conséquence immédiate qui semble sur le coup plus satisfaisante peut servir de principe à l'explication de certains types de procrastination et d'akrasie. Il serait nécessaire d'approfondir l'analyse

entrer dans le domaine de la procrastination.

Le degré d'attention temporelle lors d'une prise de décision est connecté de façon directe au contenu du processus décisionnel¹ et au type d'information traitée. C'est-à-dire que plus le processus décisionnel va nécessiter de ressources et plus le SRD va être attentif au temps actuel ou potentiel que peut prendre chaque étape. Ce qu'il est important de remarquer ici c'est que à partir d'une certaine durée du processus décisionnel – qu'il faudrait déterminer plus précisément et de façon expérimentale – le SRD prend en compte le temps comme une contrainte. Le phénomène se produit notamment lorsque le temps effectif de la prise de décision dépasse le temps que le SRD avait anticipé. Par ailleurs, le degré d'attention temporelle est aussi lié à la qualité de l'information traitée dans le processus décisionnel, par exemple si il s'agit d'un contenu représentatif concret, le temps aura l'air de passer plus vite, et le phénomène inverse se produira lorsqu'il s'agit d'un contenu représentatif abstrait².

b – La pondération du temps : délai, procrastination, et affectivité.

La TP a pour mérite indéniable de formaliser comment les sujets pondèrent les probabilités grâce à l'introduction du coefficient de pondération π , par ailleurs nous avons vu dans notre partie précédente que la pondération accompagnait l'édition des perspectives et qu'elle devait être ramenée au travail du SRD. Il est logique, dans la continuité de notre réflexion de nous interroger sur la façon dont le SRD pondère le temps et quels sont les impacts de cette pondération sur la prise de décision. Comme nous allons le voir, la pondération peut provenir de sources diverses, et produire différents effets.

Les sujets font souvent montre d'une forme d'impatience³ quant aux conséquences, aux rétributions, aux buts, ce qui nous indique qu'une forme pondération est à l'œuvre, puisque l'impatience est le signe d'une représentation distordue de la durée : le temps qui sépare le sujet de son but semble plus long que ce qu'il est réellement, par ailleurs la représentation de la durée qui sépare la décision de sa conséquence transforme la représentation de la conséquence elle-même. Le phénomène de l'*impatience* nous indique

de l'impatience et de se demander si – du point de vue des conséquences et des rétributions – l'impatience n'est pas la caractéristique essentielle du rapport qu'entretient le SRD avec les conséquences de la décision.

¹ Voir, Hansen/Trope (2012), p 8-9.

² Ibid.

³ Voir Ainslie (2005), p 636 : « There is extensive evidence that both people and nonhuman animals spontaneously value future events in inverse proportion to their expected delays ».

qu'il y a deux limites temporelles qui sont prises en considération de façon originale : l'immédiat et le délai. L'impatience nous montre aussi que le temps lui-même intervient comme élément pondérateur dans la prise de décision : si la représentation de la conséquence d'une action change selon qu'elle se situe dans un avenir plus ou moins proche, la prise de décision elle-même peut alors se trouver modifiée¹.

Dans le cours de notre réflexion, deux questions vont nous guider de façon analogique : existe-t-il un coefficient de pondération propre au temps, de type $\tau(t)$? Par ailleurs quelles sont les analogies entre la pondération des probabilités et la pondération du temps². Nous verrons que structurellement certaines comparaisons peuvent être établies, notamment à partir des travaux de Droit-Volet/Meck (2007). Nous allons étudier trois problématiques qui nous semblent importantes pour comprendre le rôle que joue le temps dans le fonctionnement du SRD: le concept de délai, le problème de la procrastination, et enfin la façon dont la perception de l'altérité intervient dans la perception du temps.

+ La contrainte de temps.

La prise en compte du temps comme facteur de soutien ou comme problème est totalement liée à l'établissement de la visée dont il peut d'ailleurs constituer un aspect essentiel. La dimension temporelle de l'environnement décisionnel conditionne par ailleurs quelles heuristiques ou stratégies vont être convoquées lors de la prise de décision³. Le schéma est simple : si lors de l'établissement de la visée une contrainte de temps intervient, cela va

¹ Il conviendrait d'étudier de façon précise comment la représentation de l'étalement des conséquences dans le temps peut avoir un effet sur la prise de décision. Mais pour cela il faudrait parvenir à établir une échelle temporelle dont les limites seraient les concepts d'immédiateté et de délai long.

² En fait des analogies ont été établies dans le cadre de la « discount utility theory » qui propose d'introduire dans la fonction d'utilité une « discount function », fonction qui joue le même rôle en contexte dynamique que le coefficient de pondération dans la théorie de la perspective. La « discount function » formalise le fait que plus les résultats sont éloignés dans le temps plus ils sont faiblement pondérés. Cela permet de formaliser l'*impatience* des sujets à partir du moment où le coefficient $F(d)$ est une fonction décroissante du délai d . Du point de vue formel la fonction d'utilité a la forme suivante, avec $F(d)$ comme discount function, $F(0)=1$, $c(d)$ formalisant la consommation de la ressource au délai d et $u(c(d))$ représentant l'utilité

immédiate de cette consommation, on a : $\sum_{d=0}^n F(d)u(c(d))$. Le problème avec cette formalisation, c'est qu'elle demeure dans le

cadre du contexte économique et respecte l'hypothèse de maximisation, elle a toutefois le mérite majeur de parvenir à formaliser le concept d'impatience dont nous parlions plus haut. Pour une analyse exhaustive des problèmes de choix dynamiques ou intertemporels, voir Daniel Read, in Koehler/Harvey (Eds.) (2004), p 424-443. Il est important de remarquer aussi que ce modèle est inspiré par la TP.

³ Comme le montrent Ariely/Zakay (2001) et Huber/Kunz (2007), la contrainte de temps peut avoir un impact sur tous les mécanismes décisionnels, en particulier dans la collecte de l'information, ce qui ressort de ces études c'est que la contrainte de temps favorisera le choix des alternatives risquées.

déterminer quel type d'heuristique ou de stratégie doit être utilisée par le SRD¹, cette opération a une influence sur la phase d'édition et sur la phase d'évaluation. Dans le cadre de situations inhabituelles ou exceptionnelles il faudra que le sujet produise une stratégie originale à partir de son propre fonds.

La contrainte de temps intervient lorsque dans l'établissement de la visée par le SRD se trouve inclus un délai que le sujet se pose lui-même ou qui lui est imposé de l'extérieur, la différence est importante, surtout dans le cadre de la procrastination comme nous le verrons plus loin. La contrainte de temps peut intervenir dans des décisions habituelles ou inhabituelles, critiques ou non, ainsi que dans les situations exceptionnelles. Contrairement à ce que l'on pourrait croire, une contrainte de temps n'amène pas toujours une baisse dans la qualité du processus décisionnel², les effets du délai sont assez nuancés. La plupart des articles qui étudient le rapport du temps à la décision le font au travers d'expériences lors desquelles les sujets doivent performer des opérations décisionnelles sous diverses contraintes de temps³.

Le délai, dont une des caractéristiques principales est la variabilité, doit être intégré au processus décisionnel comme un élément à part entière, car il a un impact sur le fonctionnement du SRD et sur les ressources qui doivent être sollicitées par ce dernier. Les conséquences de l'imposition d'un délai court sur une décision peuvent être : une discrimination de l'information mal faite, ainsi qu'un codage et une évaluation des alternatives incomplète. Auquel cas lorsque la contrainte est sévère il y a même des alternatives qui peuvent échapper à la vigilance du SRD. Cependant il ne faut pas oublier que, comme nous l'avons vu avec Gigerenzer/Brighton (2009) et avec Hwang (1994), la contrainte de temps peut aussi amener à des performances plus réussies dans certaines limites, toutefois, comme le remarque Hwang (1994) : « Increasing time pressure leads to better performance up to a certain point ; beyond that point more time pressure reduces, rather than increases, performance »⁴. Encore une fois ici, l'essentiel sera une question de configuration du système de référence général : il y aura des sujets que la contrainte de temps va motiver, et d'autres que la contrainte de temps va démotiver en les plaçant dans une situation de stress négatif.

Une contrainte de temps peut avoir des implications sur les stratégies décisionnelles

¹ Nous explorerons le fonctionnement heuristique du SRD en profondeur dans notre prochaine partie.

² C'est que met notamment avant Hwang (1994).

³ Voir, Hwang (1994), Maule/Hockey/Bdzola (2000), Ariely/Zakay (2001), Gonzalez (2005), Huber/Kunz (2007).

⁴ Voir, Hwang (1994), p 198.

utilisées par le SRD ainsi que sur la qualité de l'information : ce seront les heuristiques les plus simples et les plus efficaces qui seront convoquées¹, et l'information sera peut-être plus vague, ou de qualité différente. Il appert que sans contrainte de temps ce sont les conséquences positives qui sont le plus analysées, alors que sous contrainte de temps ce sont les conséquences négatives qui retiennent le plus l'attention². En fait nous pensons ici que si le sujet favorise la recherche des conséquences négatives c'est dans le but de pouvoir les éliminer le plus rapidement possible. Schématiquement, le sujet en contexte normal se pose la question : « qu'est-ce que je dois faire ? », alors que sous contrainte de temps il va se demander : « qu'est-ce que je ne dois pas faire ? », afin d'éliminer d'entrée les conséquences indésirables. Par ailleurs, en plus de changer les stratégies utilisées par le SRD, la contrainte de temps peut aussi modifier l'état affectif du sujet en le plongeant en contexte de stress négatif, modification qui va avoir elle aussi un impact sur le travail du SRD. Cependant encore une fois il ne faut pas rabattre trop rapidement le stress vers une contrainte négative, il peut aussi être source de stimulation comme l'ont montré Maule/Hockey/Bdzola (2000). Il y a donc une ambivalence du phénomène de délai, on ne peut pas dire de façon tranchée qu'il ait des effets strictement positifs ou négatifs³.

+ La procrastination.

C'est un phénomène qui est important dans notre travail pour deux raisons : tout d'abord, la procrastination est un fait très répandu qui vient directement et empiriquement contredire l'hypothèse de maximisation de l'utilité attendue⁴, par ailleurs la procrastination permet d'analyser comment le SRD peut échouer dans son fonctionnement normal : elle consiste à remettre à une date ou un moment ultérieurs la prise de décision ou la réalisation d'une action, et ce de façon répétée⁵. La procrastination concerne aussi bien le début de la décision ou de l'action que sa fin. Un autre intérêt de ce phénomène réside dans le fait qu'il est

¹ Le critère de temps est absolument déterminant en ce qui concerne le « choix » de l'heuristique pertinente par le SRD. Par exemple, le mimétisme : un sujet qui doit choisir son repas sous contrainte de temps dans un restaurant qu'il ne connaît pas aura tendance à imiter les autres. L'imitation fait partie des heuristiques analysées notamment par Boyd/Richardson (2005).

² Voir Huber/Kunz (2007), p 422.

³ Ce contraste se retrouve d'ailleurs dans les vues exposées par Maule/Hockey/Bdzola (2000) par rapport à celles de Huber/Kunz (2007). Contraste qui est d'ailleurs relevé par ces derniers, p 422.

⁴ Cet aspect de la procrastination a été mis en évidence par Akerlof (1991), p 1 ; et Steel (2007), p 72. L'article de Steel est fondamental en ce qui concerne l'histoire du concept, les définitions et les explications du phénomène. De son côté Ainslie (2005) met la procrastination au rang des échecs de la volonté, voir, p 635 ; voir aussi Ainslie (2010).

⁵ Voir Tappolet/Stroud (2003), p 1-2.

très commun et robuste, il peut avoir des effets dans différents domaines liés aux décisions¹. Nous allons partir d'une citation d'une citation d'Akerlof :

« Procrastination occurs when present costs are unduly salient in comparison with future costs, leading individuals to postpone tasks until tomorrow comes, the required action will be delayed yet again »².

Le premier élément à prendre en compte ici, en nous situant dans le cadre de nos hypothèses, c'est la façon dont le SRD code l'information. Comme nous l'avons vu plus haut, une des caractéristiques essentielles de ce système, c'est de maintenir les décisions habituelles des sujets autour d'un axe afin de stabiliser et de rendre cohérente l'existence subjective quant aux choix. Il se constitue ainsi un certain nombre d'habitudes décisionnelles. Dans le cadre de la procrastination, l'information que le SRD va mettre en avant, ou qu'il va surévaluer³ c'est que l'utilité de ne pas décider ou de ne pas faire ce qui doit être fait est supérieure à l'utilité de décider ou d'agir⁴. Ce mode de fonctionnement va créer une inconsistance sur le long terme, car choisir de ne pas décider lorsqu'il le faut va affecter le comportement, les décisions futures, mais aussi l'état d'esprit, par exemple de par le sentiment d'échec. Dans ce cadre-là, le SRD en posant une décision de second ordre, choisit de ne pas décider parce que l'utilité ou le plaisir pris à l'état présent pèse plus lourd que la *représentation* de l'état futur⁵. Nous avons aussi ici en germe le principe de l'akrasie, sur lequel nous reviendrons.

Comme nous l'a montré l'exemple de l'impatience, en règle générale, c'est l'immédiat

¹ Akerlof (1991), dénombre trois domaines dans lesquels le phénomène peut intervenir : dans les décisions personnelles avec ou sans deadline, dans le cas des addictions, et enfin dans le domaine de l'épargne et des retraites. De son côté Steel (2007), établit une véritable typologie que nous ne pouvons reproduire ici, mais qui comporte pas moins de quinze entrées, selon les causes et les modes de procrastination.

² Akerlof (1991), p 1. Voir aussi Steel (2007), p 66.

³ Akerlof (1991), p 2, parle de « salient information », en rapport avec le fait que les informations ou les événements qui ont plus de relief sont toujours surévalués par rapport à ceux qui en ont moins, voir aussi Ariely/Wertenbroch (2002), p 219. C'est aussi ainsi que fonctionne le SRD dans son contexte habituel, mais il peut aussi mettre lui-même en relief un certain type d'information. C'est d'ailleurs ce qui se passe dans la procrastination d'une façon négative.

⁴ Voir à ce sujet l'exemple qu'utilise Akerlof (1991), p 3-5, dont l'idée est la suivante : Akerlof était en Inde pour un an, il reçoit un ami qui oublie en repartant une valise de vêtements, chaque jour Akerlof se dit qu'il va envoyer la valise le lendemain, jusqu'au jour où il doit lui-même repartir et il amène finalement la valise de son ami avec ses affaires, car cela ne servait plus à rien de l'envoyer. L'exemple est très intéressant d'autant plus qu'Akerlof arrive à le formaliser parfaitement, voir, p 3.

⁵ Akerlof (1991), p 6-7 remarque aussi que les sujets peuvent souvent avoir seulement une représentation floue de leur propre futur, cela rejoint ce que nous avons vu concernant la prédiction des préférences. Le rôle du SRD est certes d'anticiper, mais il ne faut pas exagérer la capacité que les sujets ont de se projeter dans l'avenir de façon raisonnable, il s'agit là surtout d'une question d'éducation, car le SRD est un système qui peut être éduqué.

qui va être surévalué par rapport au délai, et cette surévaluation du présent par rapport à l'état final explique pourquoi la procrastination peut être pensée comme une nouvelle réfutation empirique de la TUA dont un des principes fondamentaux est de ne prendre en compte que les résultats, c'est-à-dire les états futurs sans les ramener à l'état présent du sujet. Par contre le fait que ce soit à partir du présent que se fait l'évaluation – même biaisée – est parfaitement en adéquation avec la TP et avec les hypothèses sur le présent du SRD que nous avons développées plus haut. C'est justement parce que le SRD n'est que présent et que le futur ne peut pas être anticipé de façon parfaitement claire que la procrastination et l'akrasie sont possibles.

Comme nous le notions plus haut, il existe des procrastinations avec ou sans délai, et dans le cadre du délai, ce dernier peut être imposé de l'extérieur ou par le sujet lui-même. Dans le cadre de la procrastination sans délai – qui est somme toute la plus rare – Le SRD code l'alternative qui peut être repoussée comme *indéfiniment repoussable*, dans ce contexte-là le sujet se dit à lui-même qu'il aura toujours le temps de réaliser l'alternative en question. Par exemple, imaginons quelqu'un qui déteste classer ses papiers administratifs, et qui les laisse s'accumuler quelque part chez lui, la tâche de classement est bien entendu codée comme une perte par rapport au fait de ne pas classer, *l'utilité immédiate c'est de ne pas classer*. Seulement il peut arriver que le sujet en question, pour une raison quelconque ait besoin à un moment donné de retrouver un papier, ce qui va lui prendre un temps considérable, et il pourra mesurer ainsi l'utilité négative de sa procrastination. Ce qui ne veut pas dire par contre que son comportement cessera une fois le papier retrouvé. Dans le cadre d'un délai imposé, son non-respect peut entraîner le « rush » que connaissent beaucoup d'étudiants. Ici les choses sont un peu différentes, car les sujets ont en général conscience que le rush va arriver à un moment ou à un autre. Mais encore une fois c'est l'état présent et sa stabilité qui vont être favorisés¹, et dans ce cas précis on ne peut pas dire que l'anticipation du futur soit obscure : l'étudiant sait que le rush va arriver, mais malgré cela le SRD le sous-évalue par rapport au présent.

¹ Steel (2007) fait de ce type de comportement l'archétype de toute procrastination, voir p 65 : « The absolute amount of procrastination is considerable, with students reporting that it typically occupies one third of their daily activities ». Steel met aussi en avant l'ancienneté du comportement que certains avaient tendance à faire émerger avec la révolution industrielle, en effet on trouve des conseils contre la procrastination dans *Les travaux et les jours* d'Hésiode. À partir du moment où existe le pouvoir de faire existe de pair le pouvoir de s'abstenir de faire : la faiblesse de la volonté émerge avec la volonté.

Dernier point, celui où c'est le sujet qui s'impose à lui-même un délai¹ pour commencer ou pour arrêter, ce cas peut se présenter comme une complexification des deux types précédents, comme l'atteinte d'une certaine limite à partir de laquelle la non-décision ou la non-action vient parasiter de façon importante la qualité du présent. Dans ce cas-là les choses doivent être un peu plus détaillées, car le processus permet de voir comment le SRD peut être déterminé par l'état d'esprit du sujet². Dans ce contexte, arrivé à un certain point de son processus de procrastination le sujet va éprouver un sentiment de honte, ou d'échec³, un sentiment d'insatisfaction lié au fait que le SRD ne parvient plus à réguler ou à contrôler le processus décisionnel de façon normale. Or, ce sentiment d'échec peut parfois produire une réorientation des phases d'édition et d'évaluation du SRD en vue de sortir du processus de procrastination⁴. Cela permet d'analyser le lien très fort qui unit le SRD aux affects, et cela permet de comprendre du même coup que pour garder le SRD loin des émotions et des affects, dans certaines situations, il faut qu'il soit éduqué dans ce sens⁵, puisque de façon spontanée il va subir l'influence des affects, influence qui ne doit pas toujours être pensée comme handicapante⁶.

Cependant la réorientation du système en fonction du sentiment d'échec peut aussi ne pas fonctionner, car nous avons vu plus haut que le SRD se stabilise autour de certaines habitudes et de certaines heuristiques qui ont un aspect positif : la cohérence de nos existences sur le court terme et le moyen terme ; mais aussi un aspect négatif : la résistance et la fixation du SRD autour de certaines habitudes qui peuvent devenir handicapantes et provoquer une inertie pathologique. Un sujet qui ne peut changer aucune de ses habitudes décisionnelles se

¹ Ce cas est étudié dans Ariely/Wertenbroch (2002).

² Dans ce cadre-là rentrent les conduites addictives de toutes sortes, où il est presque toujours question de « s'arrêter demain », voir Akerlof (1991), p 5-6, qui consacre un intéressant développement à cette question, voir aussi Ariely/Wertenbroch (2002), p 219. Dans le cadre de conduites addictives le plaisir pris à la conduite en question amène le SRD à coder tout ce qui peut s'inscrire contre elle comme une perte sèche. La surévaluation de la pratique distord tous les éléments du SRD ainsi que son fonctionnement, pourtant dans la plupart des cas les sujets ont parfaitement conscience du degré élevé de nocivité voire de morbidité lié à leur pratique. Dans le cadre des addictions sévères à la drogue par exemple, non seulement le travail du SRD est biaisé par la représentation du plaisir pris, mais le travail du SRD est complètement déterminé par la mise en œuvre des moyens pour l'obtenir. C'est ce qui se passe par exemple dans les addictions à l'héroïne ou au crack, voir Bernheim/Rangel (2009).

³ Ce sentiment peut être éveillé par la prise de conscience du sujet, mais aussi par le regard que l'autre peut poser sur son comportement sur le mode du jugement, de l'avertissement ou de la réprimande.

⁴ Il serait intéressant de se demander comment ce processus peut aussi impacter l'établissement de la visée, en remplaçant par exemple un but par un autre.

⁵ C'est ce que nous entendons quand nous disons « décider en gardant la tête froide ».

⁶ Voir notamment Zajonc (1980), Damasio (1994), Slovic et Al. (2000) et (2005). Nous reviendrons bientôt sur le lien entre les affects et les phases d'édition et d'évaluation.

trouve dans une situation pathologique qui peut avoir des conséquences dramatiques, comme c'est le cas dans les addictions¹. À partir du moment où le sentiment d'échec n'a plus de prise sur le fonctionnement du SRD et ne peut plus le réorienter nous passons du registre de la procrastination à celui de l'akrasie qui est le suivant, que résume parfaitement la formule d'Ovide : « Je vois le meilleur et je l'approuve, pourtant je fais le pire »². Car ce qui est paradoxal dans le cadre des défaillances que nous avons vues c'est que les conséquences négatives voire dramatiques sont souvent évaluées de façon claire par le SRD³. Nous reviendrons sur le problème de l'akrasie dans notre prochaine partie, mais comme nous le notions ce qui fait la différence entre l'inertie et l'akrasie, c'est la relation que le dernier comportement entretient avec l'aversion aux pertes.

+ Altérité, affectivité et temporalité.

Nous revenons maintenant sur le problème de l'altérité que nous avons déjà analysé dans le cadre de notre réflexion sur les résistances que le SRD pouvait rencontrer lors de l'établissement de la visée ou environnement décisionnel. Les décisions peuvent être orientées par la résistance métaphorique, ou pas, que peut produire autrui, ou par sa coopération, par exemple sur le mode de l'approbation ou du blâme⁴. L'idée d'introduire l'altérité dans une réflexion sur le temps décisionnel est motivée par le fait que certaines analyses ont montré que les émotions ou leur perception chez d'autres sujets pouvaient provoquer une distorsion de la perception du temps⁵. Il s'agit pour nous d'introduire la dimension affective en contexte décisionnel par le biais du concept de temps, parce que c'est une dimension importante. Les états affectifs peuvent avoir un impact sur les trois opérations du SRD que sont l'établissement de la visée, l'édition et l'évaluation.

L'idée développée par Droit-Volet/Meck à partir des travaux de Angrilli et Al. est que nos émotions, selon leurs intensités, et selon qu'elles sont positives ou négatives influencent

¹ Par ailleurs il faut noter que plus le temps passe et moins le SRD est capable d'évoluer et de s'adapter à de nouvelles situations.

² *Métamorphoses*. 7 : 20.

³ Il existe cependant un cas que met Akerlof (1991) en avant et qui résiste à ce schéma, c'est le problème des épargnes retraites, or c'est un problème intéressant en ce sens, que justement parce que les sujets n'arrivent pas à se faire une idée de leur avenir sur le long terme il repoussent le moment de mettre de l'argent de côté pour leur retraite et se retrouvent dans un état de pauvreté dès qu'ils sortent de leur vie professionnelle. Le manque de clairvoyance produit un manque de discipline vis-à-vis de l'épargne, voir Akerlof (1991), p 7.

⁴ Rappelons pour mémoire que ce type d'influence était présent dans l'article d'Allais (1952).

⁵ C'est ce que montrent les études d'Angrilli et Al. (1997) ; Droit-Volet/Meck (2007).

notre perception du temps¹. La distorsion opérée est due au degré d'excitation provoquée par l'émotion et aussi à la valeur affective attribuée par le sujet à la situation². Ce qu'il est intéressant de relever c'est que lorsque l'excitation est haute la durée d'exposition aux images négatives (cadavres, etc.) est surestimée par rapport aux images positives (scènes érotiques). D'autre part dans le contexte d'une excitation faible, la durée d'exposition aux images négatives sont sous-estimée par rapport aux images positives. Si nous exposons ces résultats sous forme de tableau nous rendons compte qu'il y a une analogie de structure entre la pondération du temps en contexte émotionnel et la pondération des probabilités en contexte de risque³ :

	<i>Excitation faible</i>	<i>Excitation haute</i>
<i>Stimuli positif</i>	Surestimation de la durée	Sous estimation de la durée
<i>Stimuli négatif</i>	Sous estimation de la durée	Surestimation de la durée

L'analogie ne doit pas cependant être poussée trop loin. Ce que nous indique ce tableau c'est que, de la même manière qu'il existe un coefficient de pondération sur les probabilités qui explique le comportement des sujets face au risque, il existe un coefficient de pondération temporel qui explique la distorsion de la perception des durées par les sujets. Et cette pondération est un facteur qui doit être pris en compte dans le fonctionnement du SRD⁴. Il s'agit de capturer ici une intuition assez commune : la distorsion subjective du temps en fonction de l'état interne du sujet et de la quantité ou de la qualité du flux informatif qui vient de l'extérieur⁵.

L'originalité de l'article de Droit-Volet/Meck réside surtout dans la démonstration de l'effet de la perception des affects d'autres individus sur le sujet⁶, cela nous permet de voir

¹ Voir, Droit-Volet/Meck (2007), p 506.

² Dans le cadre de leurs expériences, Angrilli et Al. utilisaient des images dont le contenu pouvait orienter les réactions des sujets, par exemple des images érotiques, ou d'animaux morts. Voir Angrilli et Al. (1997), p 974-977.

³ Nous renvoyons à notre partie sur la TP.

⁴ Droit-Volet/Meck (2007) ajoutent aussi que les réactions selon le degré d'excitation doivent être reliées à deux types de mécanismes distincts : « a controlled-attention mechanism for low arousal and an automatic mechanism related to motivationnal-survival systems for high arousal », p 506.

⁵ Cette distorsion permet entre autres d'expliquer les phénomènes comme l'attente et l'ennui.

⁶ L'analyse en question se trouve aux pages 507-511.

comment la perception d'autrui est intégrée au travail du SRD non seulement dans l'établissement de la visée, mais dans le cours du processus décisionnel, puisque à partir du moment où la perception des affections chez autrui modifie la perception subjective du temps, c'est l'ensemble du travail d'édition et d'évaluation qui va être orienté.

Les auteurs montrent par exemple que la vision d'un visage en colère aura tendance à accélérer la perception subjective du temps : les durées vont sembler plus longues, car un visage en colère peut tout simplement représenter une menace pour le sujet. Dans ce cadre-là, le système de référence général va passer en mode « urgence » et vigilance, et ce processus va influencer le travail du SRD, puisque ce dernier est complètement tributaire de l'état du premier. Ce qui produit la distorsion temporelle c'est la façon dont la perception de l'affect, et les représentations qu'elle éveille immédiatement, va modifier le fonctionnement de l'horloge interne du sujet.

Comme nous le remarquons, la dimension temporelle est fondamentale pour tout processus décisionnel, en ce sens qu'elle va déterminer une bonne partie de la phase d'édition (qui sera ou non complète, selon le temps disponible), et de la phase d'évaluation (sous des contraintes de temps trop importantes, l'évaluation et la comparaison des perspectives peuvent être complètement erronées). Il est temps de détailler maintenant comment fonctionne le SRD dans la phase d'édition et d'évaluation.

2) Le système de référence décisionnel et la phase d'édition.

a – **La structure logique des perspectives et des alternatives.**

Dans la TP, Kahneman et Tversky proposent de scinder les processus décisionnels en deux phases : l'édition, lors de laquelle les perspectives sont codées, et l'évaluation au cours de laquelle la perspective qui a l'utilité attendue la plus haute est choisie après avoir été calibrée par rapport à l'avoir présent du sujet et pondérée par le coefficient de pondération sur les probabilités¹. Nous avons repris ce schéma en le complexifiant, notamment en ajoutant la construction de l'environnement décisionnel par le sujet ainsi qu'en essayant de mettre en relief les composantes représentationnelles subjectives qui sont à l'œuvre dans les processus

¹ Voir Kahneman/Tversky (1979), p 28.

décisionnels, composantes qui compromettent la « calculabilité » de l'utilité attendue. Il est temps maintenant de rentrer plus en profondeur dans le travail du SRD, justement en complétant de façon aussi précise que possible l'étude des deux phases en question, étude que nous avons commencée plus haut en analysant le lien entre catégorisation et pondération.

Dans l'article de K/T l'édition et l'évaluation sont définies en fonction des problèmes proposés aux sujets dans la partie expérimentale, ce qui signifie que les perspectives sont éditées à partir d'alternatives artificielles¹. Il convient donc ici de s'interroger : comment les choses se passent-elles lorsqu'il s'agit de décisions « réelles » ? C'est-à-dire lorsque le sujet doit éditer intégralement les alternatives en fonction d'une visée qu'il peut ou non avoir construite². Mais, préalablement il convient de réfléchir au problème suivant : qu'est-ce que construire un ensemble d'alternatives en connexion avec un but donné ? À partir de quels éléments du système de référence général sont construites les alternatives ? Qu'est-ce qui déclenche la phase d'édition³ ?

Avant de commencer à élaborer ces questions, trois remarques s'imposent :

- 1) Il existe des décisions qui ne nécessitent aucun type d'édition particulier, il s'agit des décisions que nous situons à la limite de l'automatisme⁴, elles sont évaluées comme des habitudes ou des manières d'être. Dans cette catégorie se situent par exemple les réponses aux signaux routiers et l'utilisation des objets de nos environnements quotidiens, qui malgré leur caractère automatique relèvent toujours du décidable.
- 2) Le sujet n'a pas toujours clairement conscience des différentes opérations qui

¹ C'est d'ailleurs un problème que partagent la plupart des articles que nous avons étudiés et que nous étudions, il y a toujours une artificialité du protocole expérimental. Au fond qu'est-ce qui nous garantit que les sujets feraient les mêmes choix dans la réalité ? K/T étaient parfaitement au courant de ce problème, voir K/T (1979), p 19-20. Il faut donc user de prudence et de circonspection afin d'arriver justement à cerner quels peuvent être les éléments communs aux décisions en contexte expérimental et aux décisions telles que les sujets les élaborent dans la réalité. C'est peut-être ici que la philosophie doit venir compléter la psychologie cognitive notamment grâce aux outils que sont la métaphore et l'analogie.

² Comme nous l'avons vu plus haut, il y a deux façons de concevoir la visée : où elle est proposée de l'extérieur sur le mode par exemple d'un travail, d'une tâche, d'un devoir à accomplir, où elle est le fait du SRD lui-même qui pose alors le but est les moyens. Nous disions alors que dans le premier cas la décision est hétéronome quant au but, mais autonome quand aux moyens, dans le deuxième elle est autonome quant au but et aux moyens. Cependant le concept d'expertise nous permet de raffiner la deuxième caractéristique, puisque faire appel à un expert dans un domaine particulier c'est faire appel à un sujet qui va éditer les perspectives et les alternatives à notre place, et qui dans certains cas va orienter aussi l'évaluation elle-même. C'est ce qui se passe, par exemple dans le domaine médical, et dans tout domaine pouvant être soumis à une expertise.

³ Il s'agit là de question qui ne sont pas prises en comptes dans la TP de 1979 ni dans le TCP de 1992, mais elles sont abordées de façon transversales dans le volume *Choices, Values, and Frames*.

⁴ Voir notre schéma à la fin de l'introduction du chapitre, p 219.

participent à l'élaboration d'une décision. Nous séparons et dissocions des éléments qui la plupart du temps sont imbriqués et se déroulent de façon très rapide pour les décisions habituelles. Bien entendu pour des décisions inhabituelles/exceptionnelles, et/ou critiques nécessitant plus de ressources, voire même la création de nouvelles stratégies, les choses sont différentes¹.

- 3) Il existe deux types de décisions que nous appellerons « monochoix » et « polychoix », dans le premier cas de figure le SRD opère un seul choix entre plusieurs alternatives, et ce choix détermine l'ensemble de la décision. Dans le deuxième cas, le SRD doit opérer plusieurs choix entre plusieurs ensembles d'alternatives distinctes, et ce sont ces choix qui sont constitutifs de la décision. Les opérations d'édition et d'évaluation que nous allons décrire sont les mêmes pour les deux types de décisions.

Comme nous l'avons vu plus haut, le SRD est un sous-système du système de référence général, il en dépend donc quant à son état présent, état qui est fondamental puisque c'est à partir de lui que les perspectives vont être codées comme des gains ou des pertes. La première opération du SRD c'est l'établissement de la visée – la construction de l'environnement décisionnel – pour un choix particulier. La visée détermine l'espace métaphorique informatif dans lequel la décision peut avoir lieu et cela en fonction du but à atteindre, elle détermine aussi si oui ou non il y a contrainte de temps. Enfin elle commande et déclenche directement la première opération de la phase d'édition : la détermination du type de décision auquel le sujet à affaire : habituelle non-critique, habituelle critique, inhabituelle critique, inhabituelle non-critique, exceptionnelle. C'est par ailleurs aussi dans l'établissement de la visée par le SRD que vont émerger les résistances que nous avons analysées plus haut. Dernier point, mais il est capital : dans l'établissement de la visée intervient aussi – puisque son établissement est directement connecté à l'état du système de référence général – l'humeur du décideur et la façon dont il est affecté à ce moment précis².

¹ C'est ce qui ressort par exemple de l'article de Gonzalez (2005), ainsi que de Slovic et Al. (2005), mais aussi des ouvrages qui ont pour vocation d'« éduquer » ou de « coacher » la faculté de décider en détaillant et en explicitant toutes les étapes et tous les processus, voir par exemple Hammond/Keeney/Raiffa (1999), et Gilboa (2011).

² La littérature sur le rôle des affects et de l'humeur dans la prise de décision est très importante depuis Zajonc (1980), on peut entre autres se référer à l'article classique de Schwarz (2002), Loewenstein/Lerner (2003), Yuen/Lee (2003), Slovic et Al. (2000), Slovic et Al. (2005), et enfin à la très intéressante tentative d'introduire les affects dans la formalisation de la TUA par

Nous voyons donc qu'il y a trois éléments qui vont intervenir, et qui font en quelque sorte la jonction entre la visée et l'édition : l'habitude, la connaissance du contexte, et ce que d'une façon générale nous pourrions appeler l'état d'esprit du décideur, son état affectif. Ce sont ces trois dimensions et la contrainte de temps qui vont déterminer de façon complète la suite des opérations, à savoir l'édition et l'évaluation, car c'est toujours vers le connu et l'habituel que le SRD va s'orienter en premier lieu pour construire les alternatives, et ce même dans le cadre de situations non-habituelles ou exceptionnelles. Ces trois éléments donnent à tout le processus son élan et sa direction. Il est bien entendu que dans les cas de l'édition de décisions habituelles non critiques ou critiques, la mémoire va jouer un rôle majeur, notamment la mémoire affective.

La question qu'il convient alors de se poser, et que K/T n'avaient pas abordés, est la suivante : comment la visée enclenche-t-elle la phase d'édition ? Le mécanisme est assez simple : à partir du moment où la visée produit une représentation de l'espace dans lequel le but peut être atteint, elle produit de façon parallèle une *intention* qui va accompagner l'édition et l'évaluation jusqu'à la prise de décision. Cependant il faut bien prendre garde que ni l'établissement du but, ni l'intention ne sont à elles seules des garanties quant à l'optimalité de la prise de décision ou quant à la réalisation de l'action, c'est ce que nous montrent les phénomènes de la procrastination que nous avons analysés plus haut, et de l'akrasie que nous allons étudier sous peu. L'intention garantit seulement la continuité des processus décisionnels qui vont de la visée à l'édition.

La relation visée-édition est une relation qui pourrait être formalisée et cette formalisation peut nous éclairer sur la façon dont les deux concepts sont liés : soit \mathbf{B} l'ensemble des buts possibles, et B un but particulier, soit \mathbf{V} l'ensemble des visées possibles et V une visée particulière. Enfin, soit \mathbf{A} l'ensemble des alternatives possibles et a_1, a_2, \dots, a_n des alternatives particulières, éléments de \mathbf{A} . Pour tout $B \in \mathbf{B}$, il existe un $V \in \mathbf{V}$, et un sous-ensemble de $\mathbf{A} \supset \mathbf{A}$ composé d'éléments disjoints $a_1 \vee a_2 \vee \dots a_n$, tels que : $[(B \supset V) \supset (V \supset A)]$ où $A = a_1 \vee a_2 \vee \dots a_n$. D'où on peut conclure que $(B \supset A)$ ce qui signifie que la visée a pour fonction de connecter le but aux alternatives qui sont conçues comme ses

Hermalin/Isen (2006). La mise au point sur le rôle positif et constructif des affects dans la cognition a été faite par Damasio (1994), nous pourrions dire que c'est Spinoza qui l'emporte contre Descartes. Nous reviendrons sous peu sur ce problème.

possibilités de réalisation.

À partir du moment où la relation logique entre le but, la visée et l'édition des alternatives a été éclaircie, il convient de définir de façon plus exacte ce que nous devons entendre par alternatives et par perspectives dans le contexte de la phase d'édition. Une alternative peut être définie comme *la représentation d'une possibilité ou d'un moyen de produire un état du monde qui soit en adéquation avec le but que le SRD a délimité dans l'opération de visée*. L'alternative c'est l'élément, construit ou pas par le sujet, qui permet la réalisation du but. Lorsque le sujet explicite la phase d'édition dans son esprit, les alternatives ont une forme propositionnelle particulière, il s'agit de conditionnels disjoints, car il est de l'essence de l'alternative d'*impliquer* un but. La forme disjonctive exclusive est due au fait que même si il existe un nombre indéfini d'alternatives, qui sont des états du monde possibles, un seul se réalisera dans le monde actuel à l'exclusion de tous les autres. Quelques précisions doivent être apportées ici.

Dans la TP et la TCP les alternatives sont considérées comme les éléments d'une totalité plus large : la perspective, qui peut être régulière ou pas¹. Une perspective est représentée sur le mode d'un ticket de loterie, avec deux résultats pondérés par des probabilités. La perspective est elle-même pondérée en fonction de l'avoir présent du sujet et du coefficient de pondération sur les probabilités. Bien que Kahneman et Tversky ne se soient jamais expliqués de façon approfondie sur le concept de « perspective » (« prospect »)², nous pouvons faire l'hypothèse qu'il est lié à l'idée de possibilité et d'état du monde possible. Dans le cadre de la TP, la perspective reflète la possibilité – pondérée par le sujet – des alternatives qui la constituent. Par ailleurs, les perspectives sont proposées au sujet de façon extérieure sur le mode d'un questionnaire, où il doit choisir entre deux alternatives.

Nous avons ici une première distinction à opérer : de la même façon qu'il existe des buts construits par le sujet et des buts imposés³, il existe des alternatives construites par le sujet ou proposées au sujet. Par exemple si le sujet va au restaurant dans le but de s'alimenter, les choix possibles de repas seront les alternatives qui lui sont proposées de l'extérieur pour parvenir à son but. Dans ce cadre-là, les alternatives seront éditées en fonction du goût du sujet

¹ Voir Kahneman/Tversky (1979), p 30-31, ainsi que notre analyse de ces concepts.

² Il fut un temps question pour K/T de désigner leur théorie sous le concept de « value theory ».

³ C'est ce qui se passe dans la TP, mais aussi dans la plupart des expériences en psychologie où il est question de décision et de choix rationnel.

mais aussi en fonction du temps qu'il a pour opérer son choix. Si le sujet est chez lui et décide de se préparer un repas, il devra construire les alternatives à partir par exemple de ses envies et de son régime alimentaire s'il en suit un. Une deuxième remarque concerne le nombre des alternatives qui composent une perspective : il peut être supérieur à deux, c'est ce que montre la TCP en généralisant la TP¹. Le problème avec ces deux théories c'est qu'elles ne formalisent que les cas où les alternatives sont proposées au sujet de l'extérieur sur le mode du questionnaire. Or, il convient de s'interroger aussi sur la façon dont le sujet édit les alternatives et les perspectives à partir de son propre fonds, c'est-à-dire de la structure présente du SRG, et cela en fonction du but et de la fin qu'il poursuit.

Avant de passer au problème de l'influence du SRG sur l'édition des alternatives, clarifions les relations entre les concepts de perspective, alternative et but. Si nous devons synthétiser ces relations nous pourrions dire que les alternatives sont les moyens grâce auxquels le sujet peut atteindre son but, elles contiennent en elles la possibilité de modifier l'état du monde en adéquation avec la visée. Pour un but donné, deux cas peuvent se présenter : l'alternative qui va permettre sa réalisation est unique ou plusieurs alternatives sont possibles entre lesquelles le SRD, via l'évaluation opérera une discrimination². Dans ce cadre-là, une perspective sera un ensemble d'alternatives disjointes en relation avec un but particulier, une perspective sera donc la représentation des *possibilités* que le SRD a à sa disposition pour réaliser son but. Nous dirons alors que les alternatives sont homogènes quant au but. Cependant il existe un autre cas de figure, qui correspond aux exemples analysés par la TP et la TCP et dans lequel les alternatives qui constituent la perspective peuvent produire des résultats différents³, auquel cas nous dirons ici que les alternatives ne sont pas homogènes quant au but. Ce qui dans ce cas va donner une homogénéité aux perspectives c'est le calcul de la valeur combinée et pondérée des alternatives, à savoir l'utilité attendue de la perspective.

Comme nous le suggérons dans l'introduction du chapitre, la phase d'édition est elle-même scindée en deux moments : la *catégorisation* qui va permettre de déterminer à quel type de décision le SRD a affaire, comme nous le savons il y en a cinq : habituel non-critique, habituel critique, inhabituel non-critique, inhabituel critique, exceptionnel. Le deuxième

¹ Voir Kahneman/Tversky (1992), p 46-49.

² Ce qui ne veut pas dire que le SRD n'évaluera les alternatives uniques.

³ C'est le cas par exemple des alternatives qui constituent les deux options du paradoxe d'Allais.

moment de l'édition – la *contextualisation* – aura pour vocation de déterminer en fonction de la catégorisation si les alternatives peuvent être éditées dans un contexte subjectif connu ou pas. Il s'agira donc ici de déterminer quels éléments du système de référence général peuvent être sollicités sur le mode informatif pour faire face à la situation, nous verrons que dès que nous sortons des décisions habituelles les choses se compliquent. Dans cette problématique-là, c'est la possibilité de faire appel à des éléments mémorisés qui distinguent les deux types de contextualisation : en contexte connu la mémoire va intervenir massivement, en contexte inconnu beaucoup moins, voire pas du tout.

b – La construction des alternatives : le codage affectif et le rôle des croyances.

Dans le cadre de la TUA nous ne pouvons pas dire à proprement parler qu'il existe une phase d'édition qui soit explicite : l'édition se confond avec l'opération de calcul de l'utilité attendue. Autrement dit nous avons ici affaire à une édition minimale d'ordre strictement computationnelle, qui a pour but de permettre l'évaluation à partir de l'utilité espérée des alternatives. Ainsi est totalement évacué l'aspect représentationnel du décidable, ou du moins est-il réduit au minimum, d'autre part est évacué la complexité du processus de délibération. De son côté, la TP tente de combler cette lacune en introduisant dans le processus décisionnel une phase d'édition qui s'articule autour de cinq opérations : le codage, la combinaison, la ségrégation, l'annulation, la simplification et la détection de la dominance¹. Comme le remarquent les auteurs, l'édition est une *reformulation* qui a pour vocation de simplifier la représentation des perspectives. Ces opérations peuvent être considérées comme des *heuristiques d'édition*, c'est-à-dire des « méthodes » qui sont convoquées de façon quasi automatique, et autant qu'il est possible. Les effets des opérations peuvent se combiner, s'annuler, etc. Cette phase d'édition permet de penser de façon formelle comment le sujet construit les perspectives en les éditant, nous avons donc ici une des deux définitions de l'effet de cadrage (« framing effect »)² : de par les opérations de la phase d'édition, le sujet donne à la perspective une certaine forme.

¹ Voir Kahneman/Tversky (1979), p 28-29. Nous ne revenons pas ici sur les caractéristiques de chaque opération.

² Voir Kahneman/Tversky (1986). Il existe en effet deux types d'effets de cadrage : lorsque la forme de la décision change et que les alternatives restent les mêmes, elle est en partie explorée dans l'article de 1979, et totalement explicitée dans celui de 1986. Par ailleurs il existe aussi un effet de cadrage subjectif qui correspond à la façon dont le sujet construit et édite ses

Comme nous le remarquons plus haut, dans le cadre de décisions habituelles, la mémoire va intervenir de façon assez importante¹. C'est ici notamment que nous retrouvons l'aspect stabilisateur du SRD qui va maintenir l'ensemble des décisions usuelles dans un registre d'habitudes qui peuvent évoluer selon certaines règles. La principale règle étant *qu'un changement d'habitude dans le registre décisionnel quotidien doit être motivé par une augmentation du degré d'aisance cognitive et de la puissance d'agir : moins de moyens pour plus de résultats, ou moins de moyens pour les mêmes résultats*. Le problème majeur que nous rencontrons ici, et qui rend difficile l'application directe du concept d'édition de la TP dans notre théorie, c'est que les opérations de la phase d'édition – hormis le codage – ne sont pertinentes que si nous avons dans les alternatives une information probabiliste objective fiable. Cela nous amène à nous demander : lorsque nous sortons du contexte probabiliste, ces opérations peuvent-elles garder une pertinence ? La réponse est ici nuancée : les opérations gardent leur pertinence de façon analogique et métaphorique, car il est clair que la plupart du temps les sujets vont tenter de construire une représentation simplifiée des alternatives qui sont à leur disposition. Seul le codage en termes de gains et de pertes par rapport à l'état présent du système de référence général fonctionne correctement. L'édition aura donc pour fonction, en plus de catégoriser et de contextualiser la décision, de la simplifier. Mais s'arrêter là dans l'analyse de la phase d'édition ce serait ignorer que deux autres composantes entrent en jeu dans l'édition des alternatives et des perspectives : d'une part l'humeur du décideur, ses affects et ses émotions, d'autre part, à un degré qu'il va falloir réajuster², les croyances. Nous parlerons alors d'*édition affective* et d'*édition doxique*, qui sont deux formes de codages fondamentaux et déterminants dans toute prise de décision.

Ce sont aussi comme nous le verrons en conclusion deux moments où peut être introduit le concept de liberté qui n'aura été que très peu abordé dans ce travail. Il conviendra alors de montrer comment le concept de liberté peut être construit dans *le rapport que le sujet entretient avec ses représentations*, notamment ses affects et ses croyances. Nous allons retrouver une position assez proche de celle des Stoïciens.

décisions. C'est précisément ce que nous étudions ici, puisque ce cadrage est dû à l'intervention de différents facteurs représentationnels que nous détaillons. L'édition et le cadrage se produisent dans tous les cas de figure : que le sujet construise les alternatives, ou qu'elles lui soient proposées de l'extérieur.

¹ Ce type de situation est assez proche de que Goldstein/Gigerenzer (2002) entendent par « recognition heuristic »

² Voir Engel (1997), p 339.

+ Le codage affectif des perspectives et des alternatives.

Le rapport entre les affects et la décision hante la philosophie depuis Platon et Aristote¹, le plus souvent au détriment de ces derniers qui sont la plupart du temps conçu comme ce qui résiste à la détermination du meilleur par la raison. Du fait du lien des affects et des émotions avec la sensibilité et par la même avec le corps du sujet, il était du devoir de l'homme rationnel de les atténuer ou de les éliminer². Dans la philosophie occidentale il faudra attendre Spinoza pour que soit prise en compte la puissance positive des affects, par exemple de la joie qui peut augmenter la puissance d'agir du sujet, et par là même son bonheur³. D'un autre point de vue, Nietzsche reconduira les valeurs chrétiennes à leurs genèses affectives, montrant ainsi par là que la plupart des commandements moraux sont des effets des forces réactives qui ont détourné les affects de leur puissance affirmative originale pour en faire des valeurs⁴. La tradition philosophique occidentale a peut-être eu le tort de subsumer toutes les formes de décisions sous le registre de la décision morale, ainsi que d'associer de façon trop étroite la volonté avec la raison à la façon de Platon, des stoïciens ou de Kant par exemple.

Dans le cadre de la TUA, affects, émotions et humeur du décideur, ne sont tout simplement pas pris en compte du tout, ce qui est inadmissible si nous entendons passer à une philosophie de la volonté qui soit réaliste, c'est-à-dire qui tente de rendre compte de tous les éléments représentatifs entrant en ligne de compte dans les choix rationnels. Cependant il serait malhonnête d'ignorer complètement les tentatives qui ont été proposées pour formaliser certains éléments du domaine affectif dans le cadre de la TUA comme l'ont fait Loomes et Sugden dans la théorie du regret de 1982⁵, ou plus récemment Hermalin et Isen en 2006 en

¹ Probablement depuis une date antérieure, nous pensons ici en particulier à Pythagore, mais aussi à Parménide, deux penseurs qui ont particulièrement influencé Platon. Chez ce dernier l'exclusion des affects du décidable est à relié à un rejet global du corps dont témoigne par exemple le *Phédon*, le corps est ce qui doit être maîtrisé et éduqué de sorte qu'il n'entrave pas l'exercice de la raison. C'est ce qui ressort des thèses sur l'éducation physique exposées par Platon, par exemple dans la *République* ou les *Lois*. Aristote, bien que plus modéré en ce qui concerne le corps et l'utilité des plaisirs pour la vie heureuse, fera tout de même des affects l'objet d'une « métriopathie » exposée dans l'*Éthique à Nicomaque* et dans les *Politiques*.

² C'est une position commune, entre autres, aux Stoïciens, à Descartes et à Kant, ainsi qu'à Schopenhauer. C'est ce dont témoigne par exemple le pessimisme de ce dernier auteur concernant l'Amour, dans le chapitre XLIV du *Monde comme Volonté et comme Représentation*, dans lequel le sentiment en question bien loin de se voir promu au rang d'un affect pouvant stimuler l'existence est considéré comme le « Sens de l'espèce » (« Sinn der gattung »). Pour un point de vue plus contemporain sur le rôle des affects dans le décidable, voir Loewenstein/Lerner (2003), p 633.

³ Voir Spinoza, *Éthique*, Livre III, scolie de la proposition LIX.

⁴ Voir par exemple *Généalogie de la morale*, en particulier le deuxième traité. Dans le cadre de Nietzsche la valeur est un affect qui s'est retourné contre lui-même.

⁵ Voir Loomes/Sugden (1982). Il s'agit notamment dans cet article de résoudre les difficultés mises en avant dans la TP de 1979 en restant dans le cadre strict de la TUA, et en affaiblissant certains axiomes comme ceux de transitivité et

introduisant dans le formalisme de la TUA une fonction qui représente l'humeur présente ou l'affect présent¹.

Il faut noter aussi que l'intervention d'éléments subjectifs dans les décisions avait été suggérée par Allais (1953), cependant ce dernier parlait de dimension psychologique de façon générale et n'était pas allé jusqu'à préciser l'influence des affects, bien qu'il s'en soit approché dans son article de 1952². D'une certaine façon, la TP de 1979 représente une avancée dans le réalisme du fait de la formalisation de la fonction de valeur et du coefficient de pondération qui peuvent tous deux être reliés aux affects du décideur³.

Du point de vue de la psychologie cognitive, il faudra attendre Zajonc (1980) pour que soit reconnue la *primauté de l'affectif sur le cognitif*, affectif qui avait encore été éloigné du champ d'investigation puisque l'idée directrice était que le cognitif précède l'affectif⁴. Or il semblerait, et c'est dans ce sens que parle l'expérience quotidienne, que ce soit l'inverse⁵. Enfin, Damasio (1994) a montré de façon expérimentale que chez des sujets ayant subi des lésions cérébrales, l'altération des zones responsables des émotions produisait une baisse significative de la qualité des décisions⁶. Les éléments qui vont intervenir dans le travail du SRD sont les émotions et l'humeur du décideur que nous subsumons sous le concept générique d'affects, en nous inspirant de Stickney (2009) :

« Affect is a generic label that encompasses both moods and emotions. Moods are low-intensity, relatively enduring feelings usually without a known antecedent, while emotions are more intense, short lived, and have a clear cause. Both moods and emotions

d'ordonnement, et en ajoutant une fonction de désappointement dans l'équation originelle. Pour une introduction générale à la théorie du regret, voir Loewenstein/Lerner (2003), p 623-624.

¹ Voir Hermalin/Isen (2006), p 4. Ce qui est remarquable dans cette entreprise c'est que les auteurs ne *modifient pas* les axiomes de la TUA, ce qui est un tour de force car la plupart du temps lorsqu'il s'agit d'intégrer des données de type subjectif comme le regret ou l'aversion au risque, les auteurs modifient les axiomes, par exemple en les affaiblissant.

² Voir Allais (1953) et surtout (1952), p 57-60.

³ Le lien est notamment fait par Rottenstreich/Hsee (2001), p 186, et Loewenstein/Lerner (2003), p 624-625.

⁴ Voir Zajonc (1980), p 154, ainsi que Slovic et Al. (2005), p 36, qui se situent d'ailleurs dans la lignée de Zajonc.

⁵ Voir Schwarz (2002), dans Gilovitch/Griffin/Kahneman (Eds.) (2002).

⁶ Voir Damasio (1994). L'auteur montrait ainsi que les émotions ne court-circuitent pas les processus décisionnels comme l'a trop longtemps affirmé la tradition philosophique occidentale, mais au contraire qu'elles participaient de leur constitution et du bon fonctionnement de la « rationalité ». Intégrer l'émotionnel dans le rationnel et l'action rationnelle est un retournement qui doit être salué, c'est donner raison à Spinoza et Nietzsche contre Kant, Descartes, Aristote et Platon, Aristote étant le plus nuancé des quatre. C'est dans le prolongement des analyses de Zajonc et Damasio pour la psychologie, et de celles de Spinoza et Nietzsche pour la philosophie que nous allons situer nos propres réflexions. Mais là encore il faudra nuancer, car il existe des états affectifs qui court-circuitent complètement le fonctionnement du SRD, il faut donc arriver à penser un équilibre du rapport du SRD au domaine affectif.

are relevant to decision processes (...) because they have the potential to influence judgment and decision making »¹.

Il faut noter que l'humeur et l'état affectif dans lequel le sujet se trouve au moment de la décision sont déterminants pour tout le processus, et ce dans plusieurs dimensions. Par exemple certains types de décisions peuvent être élaborés en faisant usage de ce que Slovic et Al. ont nommé « The Affect Heuristic », dont l'idée est la suivante : la computation ou la délibération est remplacée par l'utilisation des affects eux-mêmes :

« Using an overall, readily available affective impression can be easier and more efficient than weighting the pros and cons of various reasons or retrieving relevant examples from memory, especially when the required judgment or decision is complex or mental resources are limited »².

Dans notre structure théorique ce type d'heuristique peut s'appliquer dans tous les cas de figure, c'est-à-dire quel que soit le type de catégorisation et le type d'édition, l'utilisation des affects peut d'ailleurs venir pallier le manque d'information lorsque nous nous situons dans des situations où le sujet n'a pas de connaissance du contexte. Mais il faut tout de même nuancer les choses : l'état affectif peut certes produire une orientation positive du travail du SRD, mais il peut aussi dans certains états affectifs intenses ou de pathologies comme les phobies produire une déviation des processus, voire même un court-circuit complet³. Par ailleurs, c'est aussi parce que les affects sont des éléments de type représentationnels qu'ils

¹ Stickney (2009), p 4 ; voir aussi Slovic et Al. (2005), p 35-36. Il est clair que du point de vue strictement philosophique cette définition pourrait être complétée par une typologie exhaustive des affects et de chacun de leurs effets sur le SRD. Mais pour cela il faudrait passer par une analyse philosophique du *Traité des passions* de Descartes, de la 4e partie de l'*Éthique* et de la troisième partie du *Traité de la nature humaine* de Hume, œuvres qui contiennent les fondements philosophiques séminaux d'une réflexion sur les affects et les passions, concepts qui seront à partir du XIXe siècle absorbés par la médecine et la psychiatrie naissante. Pour un point de vue général sur la question on peut se référer à l'ouvrage publié sous la direction de Pierre-François Moreau en 2006 : *Les passions à l'âge classique*. Pour un point de vue très intéressant sur le lien entre émotions et adaptativité, voir Cosmides/Tooby (2000), en particulier p 102-103.

² Voir Slovic et Al. (2005), p 36. Voir aussi l'article séminal pour cette problématique de Slovic et Al. (2000), dans Gilovitch/Griffin/Kahneman (Eds.) (2002), p 396-420. L'idée sous-jacente qui relaye fortement l'intuition de Zajonc (1980), est que l'ensemble des représentations subjectives que le sujet porte en lui sont relié à différents degrés avec des affects.

³ Le SRD peut être « éduqué » afin de mettre de côté l'aspect affectif qui le détermine immédiatement, c'est ce qui se passe par exemple dans l'armée, et d'une tout autre façon dans le domaine médical. Par exemple le médecin confronté à une urgence parce qu'il a été éduqué à réagir et décider de façon absolument non-affective pourra poser le geste pertinent sans crainte de se tromper, contrairement au non-médecin qui, gagné par la panique ou le dégoût, peut faire de très graves erreurs en voulant aider la personne. Pour une introduction au problème de la décision en contexte d'urgence médicale voir Resnick (2012). Par ailleurs les psychothérapies visent à éduquer ou rééduquer le SRD.

doivent être pris en compte dans notre recherche : ce sont de véritables éléments pondérateurs qui, lorsqu'ils sont présents, influencent toutes les étapes du processus décisionnel de l'établissement de la visée à l'évaluation finale. Les affects peuvent modifier l'ensemble des représentations subjectives, nous pourrions dire que les affects donnent aux représentations et à l'information une certaine « couleur » une certaine qualité, qui selon le type d'affect et l'intensité peut être ou non déterminante.

Les affects sont généralement, et au moins depuis Spinoza¹ divisés en affects positifs et négatifs, c'est un découpage qu'on pourrait dire traditionnel, et c'est celui que reprennent les psychologues cognitifs² mais aussi les économistes³. Selon que l'affect sera positif ou négatif il aura une influence particulière sur le SRD et sur ses opérations. Par exemple l'humeur du décideur pourra avoir une influence assez nette sur ses préférences et ses jugements⁴, une humeur positive entraîne des jugements de même qualité ; c'est la même chose avec les émotions que le sujet peut être amené à vivre. Dernier point important à relever, les affects ont un caractère holistique : ils pénètrent l'ensemble de la cognition, et ils peuvent être d'intensité variable. Ce qu'il est très important de retenir dans ce contexte, c'est que l'influence des affects sur le SRD nous montre à quel point ce système est imbriqué dans le système de référence général.

Nous faisons ici l'hypothèse que s'il existe des affects positifs et négatifs pouvant influencer le SRD⁵, il existe aussi un état que nous appellerons « neutre »⁶ et qui correspond à ce que le sujet vit lorsqu'il n'est soumis à aucun type d'affect. En effet il serait irréaliste de se représenter le sujet comme oscillant de façon permanente entre différents états affectifs extrêmes⁷, il faut trouver une moyenne, un état neutre qui se rapprocherait de ce que l'on entend habituellement par « humeur normale », humeur qui pourrait être pensée comme le sentiment général pris à l'existence et dont la qualité moyenne serait une certaine positivité. Cette positivité est certes problématique, et nous ne voulons pas ici tomber dans l'optimisme

¹ C'est ce qui ressort de la troisième partie de l'*Éthique* et de la dichotomie joie/tristesse.

² Voir par exemple Nygren/Isen (1996), Yuen/Lee (2003), Loewenstein/Lerner (2003), Slovic et Al. (2005), Stickney (2009).

³ Voir Hermalin/Isen (2006).

⁴ Voir Schwarz (2002), dans Gilovitch/Griffin/Kahneman (Eds.) (2002), p 534.

⁵ Des affects intenses peuvent modifier de façon temporaire l'ordonnement des préférences subjectives, mais d'une façon plus générale, la plupart des préférences subjectives sont teintées d'affect, c'est par exemple ce que mettent en avant les recherches de Slovic et Al. (2005), et (2002). Il existe donc une construction affective des préférences, et elle est importante.

⁶ Ce que Loewenstein / Lerner (2003) appellent « 'cold' state », p 629, et Yen/Lee (2003) « neutral mood », p 11.

⁷ Comme cela peut arriver dans les pathologies de type cyclothymies et troubles bipolaires.

béat, cependant elle peut et doit être postulée à partir des phénomènes que sont l'aversion au risque et les préférences marquées pour le statu quo, toutes les deux issues des recherches sur la fonction de valeur en TP¹. Comment expliquer ces phénomènes si ce n'est parce que l'état ou l'avoir présent semble avoir plus d'impact qu'une amélioration future ou probable, il semblerait, à la lumière de ces phénomènes que – *dans un contexte normal* – l'état présent ou l'avoir présent soient jugés relativement satisfaisants pour le sujet, relativement car il faut quand même pouvoir envisager une sortie du statu quo et de l'état présent, c'est-à-dire une évolution générale et conjointe du SRG et du SRD en adéquation avec le cours de l'existence.

L'impact des affects dans le travail du SRD est complexe et ne se laisse pas ramener à des alternatives parfaitement tranchées, c'est ce que met en avant une grande partie de la littérature sur la question². Nous pourrions dire que les affects interviennent dans la phase d'édition de façon ambivalente, ils peuvent être un auxiliaire ou un obstacle insurmontable, donner une orientation pertinente ou dramatique³. Un dernier point à remarquer avant de rentrer plus précisément dans l'analyse de l'intervention des affects sur l'édition est le suivant : dans cette phase ce sont les affects vécus par le sujet qui interagissent lorsqu'ils sont présents⁴, alors que dans la phase d'évaluation, en plus des affects vécus par le sujet il faudra intégrer ceux que le SRD va anticiper comme conséquences possibles de son processus décisionnel.

Il faut maintenant parvenir à préciser les effets des affects sur l'édition des alternatives et des perspectives. Nous allons reprendre ici la distinction usuelle en philosophie et en psychologie cognitive entre affects positifs et affects négatifs. Il faut d'abord remarquer que les effets des affects sont ambivalents et qu'ils dépendent du type de décision qui est en jeu : être affecté positivement dans le cadre d'une décision habituelle non critique n'a pas le même effet que dans une situation inhabituelle. Les types de décisions peuvent d'ailleurs eux-mêmes produire un certain nombre d'affects comme nous allons le voir.

Il faut arriver à restituer la complexité du phénomène. Ce qui ressort de façon générale

¹ Voir Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 143-159.

² Voir notamment Nygren et Al. (1996), p 60, et surtout p 69 ;

³ Voir à ce sujet Yuen/Lee (2003), qui partent de l'exemple médical de la dépression nerveuse, lors de laquelle les médecins conseillent souvent aux patients d'éviter des prises de décisions critiques, car l'état du SRG est tel que la totalité de l'information interne ou externe est distordue, et par là même inutilisable par le SRD.

⁴ Voir Loewenstein/Lerner (2003), p 626-633. Comme le remarquent les auteurs, les humeurs ou émotions présentent dans les sujets peuvent l'affecter directement ou indirectement et avoir des effets différents selon le type d'affect et son intensité. Nous ne pouvons malheureusement rentrer dans le détail de cette analyse.

c'est que les affects positifs peuvent produire aussi bien une recherche du risque qu'une aversion pour le risque¹. En effet un sujet qui vit un affect positif peut choisir de se maintenir dans son état, ce qui l'amènera à opérer une édition conservatrice des alternatives et à ne pas prendre de risques. Mais d'un autre côté, des affects positifs intenses peuvent conduire le sujet à négliger la recherche de l'information pertinente² et de l'information probabiliste si elle existe³. Il peut donc dans ce cas-là être amené à faire une édition partielle ou incomplète, et se laisser surprendre par des conséquences inattendues, porteuses potentielles de regret. D'autre part et pour des raisons différentes des affects négatifs peuvent entraîner les mêmes comportements : une aversion au risque et une recherche de l'information pertinente en vue de ne pas « aggraver » la situation, une recherche du risque dans le cadre d'affects négatifs intenses si le sujet a l'impression d'avoir atteint un point de non-retour, cela correspond à l'idée que quel que soit le choix la situation ne peut plus empirer. Il est important de remarquer que dans ce cadre aussi c'est par rapport à l'état présent du SRG que tout est évalué, ce qui est parfaitement en accord avec la TP⁴.

Comme nous l'avons vu plus haut l'édition des alternatives permet au sujet de leur donner la forme générale de conditionnels disjoints du type « si *a* alors *B* ». Ceci est une représentation générale, car à l'intérieur des alternatives d'autres connexions logiques sont effectives. Quand un sujet se représente le moyen qu'il peut mettre en œuvre pour atteindre sa fin – l'alternative – il pense un conditionnel, or, l'intervention des affects positifs ou négatifs peut perturber voire fausser complètement la validité de cette implication. Car l'implication en question repose sur la collecte d'une certaine quantité et d'un certain type d'information que le sujet a en lui ou qu'il doit acquérir dans le monde. Une émotion peut complètement perturber cette collecte d'information, elle peut aussi complètement déformer l'information en question. Le stress lié aux différents types de phobies est un bon exemple de la manière dont un affect peut venir altérer la qualité de l'information : le sujet sait que prendre l'ascenseur ne représente aucun danger réel, pourtant si il est claustrophobe la représentation de la possibilité de faire une crise d'angoisse va totalement changer le fonctionnement de son SRD, qui va alors chercher à produire des stratégies d'évitement. Dans ce cadre-là nous passons de

¹ Voir Nygren et Al. (1996), p 69. Voir aussi Yuen/Lee (2003), p 12.

² Voir Loewenstein/Lerner (2003), p 634.

³ Voir Slovic et Al. (2005), p 38. En fait en contexte d'influence affective, les sujets sont insensibles aux probabilités.

⁴ Cet accord avec la TP est exploré tout au long de l'article de Loewenstein/Lerner (2003).

l'implication « si je prends l'ascenseur alors je monte plus rapidement » à « si je prends l'ascenseur alors je vais faire une crise d'angoisse ». L'affect a modifié la qualité de l'implication.

Ce qui demeure le plus important dans les rapports entre la dynamique affective et le fonctionnement du SRD c'est tout d'abord, comme l'a remarqué Damasio (1994), et après lui Slovic et Al. (2003), ainsi que Loewenstein/Lerner (2003), que l'état affectif du sujet va être ce qui influence directement le fonctionnement du SRD, et augmente dans certains cas la qualité des décisions. L'affectif est premier, il oriente le cognitif¹. Par ailleurs un deuxième élément est à relever : l'ambivalence des effets des affects sur le fonctionnement du SRD met en lumière le fait que les sujets vivent leurs affects de façon totalement singulière, absolument singulière. La façon dont le sujet intégrera l'affect dans la phase d'édition sera tributaire de la forme absolument singulière de son système de référence général, c'est ce qui participera à l'originalité et à la singularité de ces décisions.

Dernier point à relever : le type de catégorie dans laquelle les alternatives vont être éditées peut lui-même produire un certain nombre d'affects. Par exemple dans le cadre de décisions habituelles critiques, le sujet peut ressentir un certain stress, dans le cadre de situations inhabituelles critiques une certaine panique face à la situation peut advenir, auquel cas le sujet peut perdre totalement ses moyens, « perdre le contrôle »², mais la situation en tant que telle peut aussi provoquer de la curiosité. Enfin dans des situations de type exceptionnel où aucun référent n'existe divers affects ou émotions peuvent survenir, comme par exemple faire le choix de se sacrifier pour le bien du plus grand nombre dans un certain contexte, et faire par là même preuve de courage.

+ Le rôle des croyances dans l'édition des perspectives.

Le système de référence général subjectif contient tout ce qui constitue le sujet comme sujet capable de représentations. Son contenu peut être actuel ou virtuel³, par ailleurs toute une

¹ Peut-être pour des raisons adaptatives ? Voir Cosmides/Tooby (2000), en particulier leur analyse de la peur aux pages 93-94. Dans ce cadre-là l'affectif oriente aussi bien le somatique que le cognitif.

² Voir Loewenstein/Lerner (2003) ; voir aussi Ainslie (2005).

³ Voir Engel (2005), p 2. Comme le souligne cet auteur il est nécessaire de postuler actualité et virtualité en ce qui concerne nos représentations, car il serait pour le moins incongru de faire l'hypothèse d'un sujet chez lequel par exemple toutes les croyances ou toutes les connaissances seraient actualisées simultanément. Si toutes nos représentations étaient explicites nous serions tout simplement perdus au milieu d'elles. Le propre d'une représentation est d'être actualisable à l'occasion d'une situation.

partie du système de référence général reste actuelle de par la présence du sujet à soi, et nous pourrions parler alors d'un sujet conscient¹. Il n'y a pas de coupure ou de rupture entre la conscience et le reste du SRG, comme nous l'avons vu plus haut, ce système en tant qu'il est réflexif est aussi le lieu des affects, affects qui peuvent avoir différents effets sur le SRD. Dans ce contexte là, le SRD est la partie du système de référence général qui prend en charge les décisions, il a aussi pour caractéristique d'être réflexif, car les décisions mêmes les plus automatiques exigent une part minimale de réflexivité².

Le contenu du système de référence général ce sont les souvenirs, les connaissances, les valeurs et les croyances du sujet, ainsi que ce qui fait l'unité et la singularité du sujet en question, à savoir d'une part le contenu représentatif propre de ces éléments et la façon dont ils sont liés, articulés et organisés, voire même hiérarchisés les uns en fonction des autres. Le SRG se définit à la fois par l'articulation absolument singulière de son contenu et par le fait que c'est un système qui peut évoluer afin de s'adapter à de nouvelles situations, ce qui veut dire que c'est un système capable d'une certaine forme d'apprentissage. Le SRD est connecté à l'ensemble des éléments représentatifs du SRG notamment, les éléments majeurs que nous signalions plus haut auxquels il faut ajouter les sentiments, la capacité de s'émouvoir, l'humeur, bref, les affects. Il est bien entendu qu'il y a une différence certaine dans la façon dont le sujet perçoit par exemple ses affects et ses croyances ou ses connaissances³. Tous les éléments que nous venons de mentionner entrent ou peuvent avoir une influence voire même une prépondérance dans le travail du SRD. Cependant nous allons nous intéresser au rôle des croyances, car de tous les éléments représentatifs sollicités par le SRD, c'est celui-ci qui a été le plus investigué par la philosophie et la psychologie.

Nous pensons que le rôle des croyances⁴ dans le processus décisionnel et en particulier

¹ Nous entendons ici ce concept dans son sens le plus élémentaire de présence à soi intentionnellement orientée.

² Cette réflexivité est aussi rendue nécessaire par le fait que le SRD doit être capable de réguler ses propres opérations et par là même de les lire. Nous faisons ici l'hypothèse, inspirée par Dennett (1978), que chaque système est auto-référent en ce sens qu'il doit contenir en lui les outils de sa propre lecture, sinon il faudrait faire intervenir d'autres éléments représentationnels et cognitifs qui auraient alors pour vocation de lire et de décoder les informations des différents systèmes, il nous semble que l'hypothèse est lourde du point de vue ontologique. Le lecteur et le décodeur doivent être intégrés au système lui-même, un peu comme dans une machine de Turing, voir à ce propos Engel (1983), p 389 et 392.

³ Ce qui ne veut pas dire que les croyances ou les connaissances ne peuvent avoir un ancrage ou une coloration affective, c'est ce qui ressort des articles de Zajonc (1980) ; Engel (1983), p 387 ; ainsi que Slovic et Al. (2000) et (2005).

⁴ C'est un concept qui préoccupe la philosophie depuis Parménide et Platon, aussi bien du point de vue éthique comme en témoignent les analyses du livre VI de *La République*, qu'épistémologique comme en témoigne le *Ménon* avec l'hypothèse de la connaissance comme croyance vraie justifiée, définition qui fera couler et fait encore couler beaucoup d'encre depuis sa « critique » par Gettier. Pour un point de vue général sur les croyances, voir Engel « Les croyances », in D. Kambouchner, (ed) (1995). *Les notions de philosophie*, Paris, Gallimard, 1995, tome II, 1-101. La littérature sur ce concept est immense il

dans la phase d'édition doit être réévalué, il convient de se poser la question suivante : s'il semble acquis par une partie de la tradition philosophique continentale – notamment aristotélicienne et kantienne – et analytique – notamment davidsonienne – que les croyances et les désirs, qui sont des représentations, ont un pouvoir causal quant à l'action¹, comment doit alors être envisagée leur influence sur les processus décisionnels que nous sommes en train de dégager puisque nos analyses se situent en amont de l'action. À l'issue de cette étude il conviendra peut-être de minorer le rôle des croyances dans la prise de décision, ou en tout cas de le nuancer. Pour ce faire il faudra d'abord opérer quelques distinctions notamment à l'intérieur du concept de croyance.

Avant toute chose il convient de s'interroger sur les liens entre la TUA, la TP et le concept de croyance. Dans leur ouvrage N/M ne tentent à aucun moment de formaliser ce concept et son rôle dans la prise de décision. Faut-il voir ici une carence comme lorsqu'il s'agissait des affects ? À notre sens ici les choses doivent être nuancées : il se pourrait que dans l'esprit de N/M, la fonction de croyance soit intégrée dans la fonction d'utilité et ce d'une façon très simple : le sujet choisit cette alternative parce qu'il *croit* qu'elle est la meilleure, cependant comme dans cette théorie nous passons par une phase de computation, la croyance en question aura certaines caractéristiques semblables avec une connaissance : le calcul permettant en quelque sorte d'objectiver la croyance. Nous pourrions alors dire que le sujet croit que cette option est la bonne puisque c'est ce que le calcul lui a permis de déterminer. Il y a aurait donc ici une strate implicite de croyance qui ne serait pas formalisée mais que refléterait le théorème de la représentation. Dans le cadre des théories de Ramsey et de Savage, la croyance joue un rôle de premier plan que nous avons analysé plus haut, elle est le concept central des deux axiomatiques qui se situent dans une perspective bayésienne, et donc dans une perspective où les croyances et leurs mesures probabilistes sont centrales². Si nous évacuons la difficulté concernant la mesure probabiliste des croyances dont nous avons déjà parlé, il faut remarquer que dans le cadre du bayésianisme ce n'est pas le degré de croyance en

est hors de propos ici d'en envisager une revue. Cependant nous conseillons vivement au lecteur de se reporter aux articles de P. Engel que nous citons, ainsi que tous ceux où il aborde le concept de croyance, car l'approche de cet auteur, en plus d'être exhaustive, est philosophiquement déterminante.

¹ Voir Davidson (1980), p 3-19, Engel (1997), p 327, et surtout Engel (2005), qui est exclusivement consacré à la question de la croyance comme disposition à l'action chez Bain, Peirce, James et Ramsey. Kant n'est certes pas un pragmatiste, mais sa position qui consiste à agir selon la représentation de la loi attribue un aspect causal à une représentation, qui n'est certes pas une croyance au sens strict, mais qui pourrait s'en rapprocher.

² Voir Oaksford/Chater (2007) et Chater/Oaksford (2008).

un fait ou un événement qui est central, mais le degré de croyance en la proposition qui le représente, et comme nous allons le voir dans un instant ce type de croyance pourrait tout aussi bien être appelé *jugement*. Il conviendra donc de distinguer entre les deux. Dernier point, qu'en est-il du rapport entre la TP et le concept de croyance ? À notre sens puisque nous avons affaire à une théorie qui passe par la computation de l'utilité attendue nous nous retrouvons dans le même cas que la théorie de N/M : les croyances doivent être intégrées de façon implicite. De plus, nous pouvons dire que d'une part dans le concept de fonction de valeur est intégré – au moins de façon minimale – les croyances du sujet quand à son avoir présent, par ailleurs l'intervention du coefficient de pondération probabiliste pourrait tout aussi être considéré comme le reflet des croyances subjectives à propos des probabilités¹. Ce qui ressortirait de ces théories, c'est qu'elles ne « formalisent » qu'un seul aspect de la croyance, à savoir celui qui se rapproche le plus du jugement.

Il existe deux façons de concevoir les croyances qui peuvent avoir différents effets sur le travail du SRD. Tout d'abord une croyance peut être considérée comme une disposition interne, un état intentionnel², c'est le cas par exemple lorsqu'un sujet croit en Dieu. Dans ce registre-là, une croyance est un engagement, une conviction, une position que prend le sujet vis-à-vis de son objet ; un engagement ou investissement qui bien que n'étant pas épistémique – voire même étant l'opposé d'une connaissance – et ne constituant pas une connaissance au sens strict, joue un rôle un peu analogue à celui d'une connaissance. Ce type de croyance a une fonction très importante dans le SRG auquel il peut même donner sa forme particulière. Bien que n'étant pas explicites ou actualisées de façon permanente ces croyances, qui peuvent être très diverses et nombreuses participent de la singularité du système de référence général³. Ces croyances sont susceptibles de degrés et d'intensités et elles peuvent donc déterminer et orienter le travail du SRD de différentes façons⁴. La deuxième façon de concevoir la croyance

¹ C'est ce que représenteraient alors aussi les heuristiques et les biais que nous avons étudiés plus haut.

² Voir Engel (1983), p 385 ; (1997), p 328-331, où l'auteur propose de fines distinctions conceptuelles pour construire la différence entre croyance et jugement ; et (2001), p 429.

³ Ce qui ne veut pas dire qu'il n'existe pas de croyances collectives ou partagées, en tout cas il existe des *objets* de croyance collective, car il est difficile de pouvoir assurer que le mode d'engagement et d'investissement des sujets vis-à-vis de l'objet soit le même.

⁴ Par exemple, si nous reprenons l'exemple de la croyance en dieu, un croyant qui a une foi mitigée et plutôt culturelle se comportera différemment et décidera différemment de celui qui aura une foi intégrale qui déterminera l'ensemble de ses décisions et de ses actions de façon drastique. L'expérience dramatique des jeunesses hitlériennes nous montre qu'avec le corps le but est de dresser la volonté, le fanatisme survient alors quand le travail du SRD est complètement dominé et borné par la croyance, politique ou religieuse. Nous retrouvons ici un schéma assez proche de celui de l'addiction, d'ailleurs nous pensons que le fanatisme est une forme d'addiction.

peut être subsumée sous le concept d'« attitude propositionnelle »¹, la croyance sera susceptible d'être vraie ou fausse de façon objective, contrairement à la croyance du premier genre auquel le sujet attribue certes des degrés d'intensité mais en la supposant vraie.

Dans le deuxième genre, la croyance se rapproche donc assez du concept de jugement², puisqu'il s'agit d'attribuer à une proposition un certain degré de vérité par rapport à un état de fait objectif, par exemple quelle est la probabilité que j'attribue à la proposition : « demain il va pleuvoir »³. C'est le type de croyances que formalisent le bayésianisme et la logique inductive. Ici la croyance est un acte, une *opération* cognitive, alors que dans le premier genre elle est plutôt un *état* oscillant entre l'activité et l'acceptation. Par ailleurs, la croyance du deuxième genre peut être plus contextuelle, alors que celle du premier est plus générale et peut durer plus longtemps, ce qui signifie que dans le deuxième genre cas une réfutation ou une modification de la croyance par l'expérience est possible, mais beaucoup moins dans le premier.

Si nous revenons à la discussion que nous avons eue plus haut sur les croyances⁴ nous nous apercevons que seules les croyances du premier genre sont soumises de façon stricte au holisme que nous postulons, puisque les croyances du deuxième type, du fait qu'elles puissent être seulement contextuelles sont susceptibles de se modifier de façon drastique, voire de disparaître complètement, leur possibilité de fluctuation rapide les empêche de participer pleinement à la configuration holistique du SRG. Cependant elles peuvent donner à ce système une texture qui bien que passagère peut intervenir dans le travail du SRD.

¹ Voir Engel (1983), p 385 ; (1997), p 330-331.

² Voir Engel (1997), p 328.

³ Voir l'intéressante discussion de Engel (1997), p 333-337, où l'auteur explore la problématique de la cohérence dynamique des croyances, et de l'engagement des sujets vis-à-vis de celles-ci.

⁴ Par ailleurs, une autre question serait de savoir comment nous acquerrons nos croyances et si ce processus est involontaire ou non. La réponse à la première partie de la question nous amènerait trop loin de notre objet, par contre nous pouvons donner des éléments de réponse pour la deuxième. Dans le premier genre de croyance l'intervention du vouloir, c'est-à-dire des processus que nous sommes en train de décrire – visée, édition, évaluation – est possible sur le registre général de la « volonté de croire », pour reprendre un titre de W. James (1897), dans ce cas-là la croyance elle-même est codée comme un but, c'est le fait d'arriver à croire qui va déterminer l'établissement de la visée et les deux phases subséquentes. Mais ce type de décision peut mener à des paradoxes structurellement semblables au paradoxe de Moore, comme le souligne Engel (2001), p 433. Cependant ce type de décision peut aussi être le reflet de ce que nous entendons généralement dans l'énoncé : « je veux y croire », il s'agit en l'occurrence de produire une forme d'auto-persuasion. Dans le cadre de cette problématique-là, il faut aussi avoir à l'esprit qu'il existe tout un ensemble de croyances tacites, qui forment le fond du SRG et ne sont convoquées que dans des circonstances très particulières, croyances dont il est difficile de retracer l'acquisition par rapport à d'autres qui peuvent être d'origine sociale, culturelle, etc. ; voir Engel (2005). Dans le cadre du deuxième type de croyances, celles qui se rapproche du jugement, et sont du registre de l'opinion, nous les acquerrons, les révisons et les perdons de façon plus courante, de façon volontaire ou involontaire, elles peuvent d'ailleurs être produites par le travail du SRD comme nous le verrons sous peu.

La façon traditionnelle¹ d'analyser l'intervention causale des croyances dans la prise de décision et dans l'explication de l'action est de les insérer dans ce que nous avons coutume d'appeler depuis Aristote un syllogisme pratique², syllogisme dont la structure est la suivante : « X désire que p, et *croit* que faire A est le meilleur moyen pour obtenir que p ; par conséquent l'agent fait A »³. Plusieurs difficultés émergent ici, tout d'abord en quel sens le terme croyance est-il employé, est-ce une croyance de premier genre ou de deuxième genre ? En fait si nous regardons les choses d'un peu plus près nous nous rendons compte que c'est le deuxième genre qui s'applique ici puisque ce que signifie implicitement le syllogisme c'est que X attribue à l'énoncé « A est le meilleur moyen pour obtenir p » un certain degré de vérité qui le pousse à agir, nous revenons ici à une certaine forme de pragmatisme que nous avons rencontré chez Ramsey ou Savage. Par ailleurs associer la prise de décision à une structure syllogistique semble à notre avis trop linéaire : c'est faire l'économie des processus qui sont à l'œuvre dans la phase d'édition et la phase d'évaluation⁴. Il faut garder à l'esprit que des facteurs comme le temps et les affects interviennent de façon première dans la construction subjective des alternatives en rajoutant ainsi des conditions et des contraintes sur le travail du SRD, et ces éléments ne peuvent pas être pensés comme un simple ensemble de prémisses qu'il faudrait intégrer à la structure syllogistique. Certes la structure est pertinente d'une façon générale puisqu'elle formalise une intuition assez courante qui dit que si le sujet vise une fin p et qu'il croit que le moyen pour l'atteindre est A, il doit faire A. Mais c'est justement dans l'édition des moyens, c'est-à-dire leur construction subjective dans l'élément de la représentation et dans leur évaluation que repose toute la difficulté. La thèse classique qui présuppose que nous décidons à partir de nos croyances et de nos désirs n'est pas fausse, elle est partielle, car elle ne rend pas compte de la complexité des processus à l'œuvre dans la prise de décision, processus que nous tentons justement de rendre le plus explicite possible. Comme le dit P. Engel, que nous suivions ici absolument :

« C'est un *dogme* de la philosophie de l'action et de l'esprit contemporaine que toute explication de l'action doit procéder en termes de croyances et de désirs. (...). Mais même

¹ Nous entendons ici aussi bien les écoles continentales qu'analytiques, voir par exemple Davidson (2001).

² Aristote expose sa théorie du syllogisme pratique dans quatre de ses ouvrages : *De Anima*, *Éthique à Nicomaque*, *Métaphysique*, et enfin *Du mouvement des animaux*.

³ Engel (1997), p 327. C'est nous qui soulignons.

⁴ Nous renvoyons à notre schéma supra, p 219.

si nous acceptons que ce schème est en général correct, il est tout à fait partiel. Il ne permet pas de donner à nos explications de l'action le *grain fin* qui est requis quand nous devons considérer le degré d'*engagement* de l'agent vis-à-vis de ses représentations »¹.

Engagement qui peut être vu de plusieurs façons : par rapport aux croyances et à leurs intensités, mais aussi dans la façon dont le sujet construit à partir de sa visée, de son propre fonds et des informations extérieures, tout le processus décisionnel. C'est ce que nous tentons de faire ici en incluant dans ce processus tous les éléments représentatifs qui peuvent nous permettre d'affiner autant que faire se peut la compréhension du décider humain. Si les croyances ont un rôle causal, ce que nous ne dénisons aucunement ici, elles ont un rôle causal qui n'est que partiel,² car comme nous sommes en train de le montrer il y a d'autres éléments qui entrent en jeu et qui ne peuvent être ramenés à des croyances. Il ne s'agit pas ici d'évacuer les croyances, seulement de minorer le monopole causal qu'elles avaient acquis dans la tradition philosophique. Les croyances peuvent déclencher le travail du SRD, le soutenir et l'orienter, mais aucunement d'une façon qui soit isolée des autres éléments représentatifs, ni du type de décision auquel le sujet à affaire.

Il est temps maintenant de nous interroger sur la façon dont les croyances peuvent intervenir dans les différents types de décisions que nous avons dégagées plus haut. Encore une fois, et comme pour les affects, pour arriver à analyser « le grain fin » de l'influence des croyances sur la phase d'édition il faudrait détailler encore plus précisément les différents types de croyances et les analyser. En fait il faudrait aller plus loin, il faudrait analyser la façon singulière dont le système de croyances est organisé chez chaque sujet, car bien que pouvant être « normalisé », par les rapports sociaux, familiaux, l'éducation, et tous les processus de structuration, il reste absolument singulier. Et c'est cette singularité que tentent par exemple de bannir l'endoctrinement et la propagande.

Dans le cadre de notre théorie les deux genres de croyances que nous avons distingués plus haut vont intervenir dans la phase d'édition. Le premier genre de croyances peut jouer deux rôles, à savoir déclencher le processus décisionnel ou le soutenir et l'orienter. Il semblerait donc que l'effet d'une croyance soit alors difficilement discernable de l'effet d'un

¹ Engel (1997), p 339. C'est nous qui soulignons.

² Il est notable que les croyances du premier type ne soient pas prises en compte dans la structure du syllogisme pratique, pourtant il est clair qu'elles peuvent tout à fait avoir un rôle et même un rôle prépondérant dans la construction des décisions.

affect. Outre le fait que les affects sont vécus et que les croyances sont conçues, la différence fondamentale réside dans le contenu informatif que véhiculent les deux phénomènes. Un affect donne au sujet une information sur ce qu'il ressent et comment il se sent face à une situation donnée, alors qu'une croyance contient des éléments informatifs qui dépassent largement celui du ressenti, bien qu'elle puisse être associée à des éléments affectifs. Les croyances du premier genre vont donc participer à la construction des alternatives dans la phase d'édition en ce sens qu'elles vont être pour le SRD une source d'information à laquelle il pourra puiser, car il faut garder à l'esprit que c'est le SRD qui sélectionne les informations pertinentes dans le SRG. Les croyances vont donner aux alternatives une partie de leur contenu sémantique, qui pourra ou devra être complété ou corrigé selon le type de décision envisagé. Complété par exemple par des souvenirs, des connaissances, l'activation de certaines valeurs, une recherche de l'information à l'extérieur. Et nous pouvons faire l'hypothèse ici que le régime doxique du SRD sera proportionnel à la catégorie dans laquelle la décision se situe, c'est-à-dire que plus la décision en question sera habituelle, routinière et quotidienne et plus le SRD ira puiser dans le fond de croyances du SRG. Et ce pour la simple raison que la plupart des habitudes décisionnelles sur lesquelles repose le travail quotidien du SRD sont elles-mêmes reliées à des croyances d'ordre général, croyance par exemple en la stabilité de nos environnements quotidiens, en la continuité de l'existence, etc. Dans ce contexte-là, les croyances d'ordre religieux, éthique, politique, écologique, etc., jouent le même rôle puisque l'investissement subjectif est constant.

Il va donc y avoir *constitution d'habitudes autour de certains éléments doxiques* qui peuvent alors devenir des pôles autour duquel le SRD va osciller. Si le sujet « croit » à une doctrine politique, cette doctrine viendra influencer toute la sphère du décidable, selon le degré de croyance qu'il investira dans la politique en question. Si il est sympathisant, l'influence sera minime, si il est militant, elle sera déjà plus importante, et orientera de façon plus décisive édition et évaluation, s'il est fanatique sa croyance aura atteint un point tel, que comme dans une addiction ce n'est plus lui qui décidera à proprement parler, mais la croyance à laquelle il se voue corps et âme. Tout, ici encore est une question de degré, et pour les

croyances il s'agira du degré d'investissement du sujet face à sa représentation¹. Les croyances du deuxième genre, celles que nous appellerons provisoirement « bayésiennes », ont une action plus localisée sur la phase d'édition, elle permettent de mesurer le degré de vérité que le sujet attribue aux alternatives qui lui sont proposées ou qu'il produit lui-même. À ce titre, elles interviennent après les croyances du premier genre. Comme nous l'avons vu plus haut les alternatives – lorsqu'il y en a plusieurs – sont éditées comme des disjonctions exclusives d'implication du type : $[(a_1 \supset B) \vee (a_2 \supset B) \vee, \dots, (a_n \supset B)]$. La croyance de genre deux consiste à attribuer un degré de vérité, situé sur une échelle subjective² aux différentes alternatives du type $(a_n \supset B)$ qui composent la perspective. Ce type de croyance va donc avoir une certaine influence sur la phase d'évaluation. Seulement il ne faudrait pas croire qu'elle puisse déclencher à elle seule l'élection de la perspective qui va être choisie, et c'est ici que rentre en jeu l'évaluation proprement dite, dont les mécanismes sont ici aussi assez complexes. Cette phase d'évaluation doit nous permettre de comprendre à la fois comment l'alternative dominante est élue, mais aussi comment il est possible que la même alternative, qui domine toutes les autres, ne soit pas choisie, comme c'est le cas par exemple dans le phénomène de l'akrasie.

3) L'évaluation et la prise de décision.

a – **Délibération, statu quo et akrasie.**

+ Édition et évaluation

Le concept d'évaluation que nous allons construire ici est directement issu de la Théorie de la Perspective et il recouvre ce que nous entendons usuellement par la *délibération*. Dans cette phase il s'agit pour le SRD de comparer entre elles les alternatives qui ont été éditées et d'élire celle qui aux yeux du sujet a la plus grande valeur en fonction du but que le sujet a établi dans sa visée. C'est dans cette phase que le décideur pèse « le pour et le contre »

¹ Il est important de garder à l'esprit que cela est vrai pour toutes les représentations que le sujet a en lui : souvenirs – qui ont une vivacité propre –, connaissances, valeurs, etc. L'influence de toutes ces représentations dans l'édition et l'évaluation ne peut être mesurée que par le degré d'investissement du sujet vis-à-vis de ses représentations.

² Comme nous l'avons montré plus haut les sujets n'utilisent pas à proprement parler une échelle de probabilités objectives. C'est ce qui ressortait des critiques de la TUA que nous avons étudiées dans notre deuxième chapitre.

comme on le dit communément. Il reste donc à analyser à partir de quoi les options sont évaluées et comment la décision en tant que telle est prise, comme nous l'avons vu dans l'introduction du chapitre il y a trois cas possibles : évaluation automatique sans édition, application d'une heuristique qui peut devenir habituelle, élaboration d'une nouvelle stratégie. Dans la TP les options sont évaluées à partir de l'avoir présent du sujet représenté par la fonction de valeur ainsi qu'en fonction de la pondération probabiliste des perspectives. Nous allons maintenant le premier point en nous rapportant non plus seulement à l'avoir présent du sujet mais à son état présent, en ce qui concerne l'évaluation en termes de pondération probabiliste nos analyses précédentes ont montré que d'une part la pondération est une opération générale liée au concept de représentation qui prend effet avant la phase d'évaluation notamment dans l'établissement de la visée et lors de l'édition, d'autre part nous savons aussi que très peu d'alternatives et de perspectives sont codées avec une mesure probabiliste. La plupart du temps, c'est la notion de « chance de » qui tient lieu d'évaluation probabiliste.

Reprenons rapidement la façon dont le SRD fonctionne en récapitulant les étapes principales de son travail. En ce qui concerne les décisions automatiques, les manières d'être ou de faire, nous ne pouvons pas à proprement parler d'établissement de visée ou d'édition qui soient explicites : le SRD actualise et mobilise directement les éléments du SRG qui sont nécessaires à la réalisation de l'objectif : le feu est rouge, le sujet appuie sur la pédale de frein. Nous sommes ici à la limite du décidable. Nous reviendrons sur ce type d'évaluation sous peu.

Pour les autres décisions le cheminement sera un peu plus complexe. L'établissement de la visée par le sujet donnera une première orientation au travail du SRD selon qu'elle met en avant un certain type de résistance, qu'un délai et présent, et que certains affects ou certaines croyances du premier genre sont en jeu. La visée ajoutée à ces éléments produira la première opération de la phase d'édition, à savoir la catégorisation selon les concepts de décision habituelle, habituelle critique, non habituelle, non habituelle critique, exceptionnelle. Cette catégorisation ajoutée aux informations déjà présentes – à savoir la visée et les éléments qui en dépendent – produira à son tour la contextualisation dans le registre du connu ou de l'inconnu, la contextualisation permet au SRD de déterminer quel rôle la mémoire et les connaissances du sujet peuvent jouer dans l'édition des alternatives, en particulier dans l'édition sémantique c'est-à-dire la procédure qui construit le sens des alternatives. L'ajout de

la contextualisation à tout le processus finalisera l'édition des alternatives, qui du point de vue de la forme ne varieront guère, en y incluant tous les éléments informatifs et affectifs convoqués pour cette décision particulière.

Le premier élément qui est pris en compte dans la phase d'évaluation c'est la *quantité*¹ et la *qualité* des alternatives éditées par le SRD. Comme nous l'avons vu plus haut, le nombre d'alternatives pertinentes pour un but donné est hautement variable. L'alternative peut être unique, auquel cas elle sera évaluée quand même, en particulier si elle est liée à des éléments affectifs, car le travail de l'évaluation c'est aussi d'anticiper les émotions liées aux conséquences de la décision. L'évaluation de ces affects potentiels, en particulier des affects négatifs, peut avoir une influence déterminante sur la prise de décision en tant que telle², le SRD pratique alors ce que Luce/Payne/Bettman (1999) appellent « Emotional Trade-Off ». Pourquoi cette opération n'est pas effectuée avant la phase d'édition ? Parce qu'il faut que les alternatives soient déjà construites pour que le SRD puisse évaluer « où » elles peuvent mener du point de vue affectif, mais aussi à tous les autres points de vue.

La qualité sémantique des alternatives, c'est-à-dire leur sens, leur contenu informatif, qui peut provenir de sources multiples, notamment de la collecte d'information, des croyances et des connaissances subjectives stockées dans la mémoire, est le résultat de la phase d'édition, qui comme nous l'avons vu élabore les alternatives du point de vue syntaxique et du point de vue du sens. En ce qui concerne la quantité des alternatives pour une décision donnée l'édition peut être exhaustive – si le SRD envisage toutes les possibilités –, ou non – si le SRD édite de façon explicite seulement une partie de l'ensemble des alternatives –, et cette exhaustivité ou sa privation peut être liée à l'intervention de divers éléments, notamment affectifs ou temporels, ou à la catégorie dans laquelle la décision se situe : critique ou pas.

Des états affectifs intenses ou des contraintes de temps jugés trop sévères peuvent altérer la phase d'édition et faire en sorte que le SRD « oublie » ou « néglige » d'éditer certaines alternatives, les mêmes états peuvent avoir des effets sur la phase d'évaluation et amener le SRD à « oublier » d'évaluer les conséquences de certaines alternatives, ce qui peut conduire au regret. Du point de vue de la qualité des alternatives encore une fois c'est une

¹ Voir Bettman/Luce/Payne (1998), p 200.

² Idem, p 205-206. Voir aussi Luce/Bettman/Payne (1997), p 386-389, et Luce/Payne/Bettman (1999), p 144-146. L'influence peut être telle qu'elle provoque une seconde édition à la lueur de cette évaluation.

question de degré, les choses pouvant aller des alternatives parfaitement conçues aux alternatives à peine pensées ou ébauchées. Il semble que plus la décision soit critique et plus les alternatives fassent l'objet d'une attention particulière, aussi bien dans le registre qualitatif que quantitatif, à l'opposé, moins la situation va être critique et moins les alternatives auront besoin d'être claires et exhaustives. C'est une règle générale, mais elle peut varier. La plupart du temps c'est le caractère critique ou non critique de la situation qui va déterminer jusqu'à quel point l'ensemble des alternatives doit être exhaustif et quelle doit être leur précision.

+ L'élimination des alternatives et les quatre règles de l'évaluation.

À la suite de l'édition, le SRD va produire une première limitation dans le nombre des alternatives, il utilise ce que nous pourrions appeler une « *heuristique d'élimination* » et élimine d'emblée les alternatives irréalistes, trop coûteuses et immédiatement dominées¹. L'ensemble du travail du SRD dans la phase d'évaluation va être conditionné par quatre règles générales qu'ont thématiqué Bettman/Luce/Payne (1998) et que nous reprenons ici, car elles s'intègrent parfaitement dans notre théorie, par ailleurs elles permettent de comprendre le lien entre la phase d'évaluation et les affects, mais aussi de comprendre dans quelle mesure le SRD peut être vu comme un *système conservateur*. Ces règles doivent néanmoins être complétées par une analyse des liens entre les concepts d'évaluation, d'avoir présent, et du couple gains/pertes, car ces concepts – rapportés au travail du SRD – permettent de proposer une explication originale d'un certain type d'akrasie et montrent de quelle manière il est relié avec un fait que la TP avait établi, à savoir que les pertes pèsent plus lourds que les gains.

La première règle que doit suivre le SRD est de *maximiser la précision de la décision*². Ce qui signifie deux choses, tout d'abord que le but doit être déterminé par et dans la visée de façon aussi précise que possible, deuxièmement, que lors de l'évaluation, le SRD sera attentif à la précision des alternatives en fonction de ce but. Nous pouvons donc postuler qu'en plus de l'élimination analysée plus haut, il y aura aussi une élimination selon le degré de précision de l'alternative. N'oublions pas que ce degré de précision est directement lié à la phase d'édition, le degré de précision d'une alternative dépend entre autres du degré de complexité de

¹ Il ne faut pas confondre cette heuristique avec l'élimination par attribut élaborée par Tversky (1972), bien qu'elle ne lui soit pas totalement étrangère puisqu'elle se fonde sur certaines caractéristiques des alternatives éditées. Nous retrouvons ici l'opération de détection de la dominance qui était intégrée à la phase d'édition de la TP.

² Voir Bettman/Luce/Payne (1998), p 199-205.

l'information qui la définit, ainsi que de la prévisibilité de ses conséquences : plus une alternative est précise et plus ses conséquences sont « lisibles ».

La deuxième règle consiste pour le SRD à *minimiser et à atténuer l'influence des émotions négatives sur la prise de décision*¹. Ce point est lié au fait que le SRD est parfois obligé d'opérer une discrimination entre différents attributs dans un ensemble d'alternatives, attributs qui peuvent faire l'objet des préférences du sujet mais que ce dernier doit sacrifier. C'est l'essence même du concept de choix en ce sens que dans un choix seulement une alternative va être élue, et ce peut être parfois au détriment d'autres alternatives qui pouvaient être considérées par le SRD comme avantageuses ou satisfaisantes aussi, mais moins importantes. C'est le cas par exemple lorsque le sujet suit un régime et qu'il doit « négocier » entre le plaisir de manger et la privation que constitue le régime, par exemple en ce qui concerne les sucreries, etc. Dans ce contexte-là, si le SRD fonctionne normalement et qu'il n'est soumis ni à la procrastination ni à l'akrasie, il va élire la perspective qui lui semble la plus satisfaisante à tous les points de vue afin de minimiser l'expérience d'émotion négative lors de la prise de décision, ainsi que l'émotion négative qu'il peut anticiper comme accompagnant la réalisation de l'alternative. Il existe des cas où ce type de « négociation » entre la satisfaction et la décision n'est pas possible, et il ne faut pas oublier qu'il existe des alternatives qui ne peuvent donner lieu à aucune négociation, car elles sont incommensurables entre elles². Enfin, comme nous l'avons vu plus haut, le fait pour le SRD d'être soumis à un affect négatif peut entraîner une édition plus longue et plus coûteuse, car dans ce contexte une attention très importante est accordée à l'information négative, ce qui peut causer un processus généralement plus long.

La troisième règle consiste à *minimiser l'effort cognitif requis pour la décision*³, c'est ici que va intervenir de nouveau la théorie de la rationalité limitée de Simon et Gigerenzer, ainsi que le concept d'heuristique décisionnelle⁴. Nous détaillerons ce problème lorsque nous aborderons le fonctionnement heuristique de la phase d'évaluation. Ce qu'il faut remarquer ici, c'est que ce ne sont pas toutes les décisions qui sont susceptibles d'obéir à cette règle. Les

¹ Idem, p 205-206.

² On trouve un exemple dramatique de ce genre de situation dans *Le choix de Sophie*, de S. Styron.

³ Idem, p 205-207.

⁴ Bettman/Luce/Payne (1998), p 187, se rapportent d'ailleurs expressément à la théorie de la rationalité limitée de Simon (1955). Il est important de garder à l'esprit que tout le travail du SRD se situe dans ce contexte.

décisions critiques impliquent souvent un investissement cognitif considérable, dont le sujet ne peut pas réellement faire l'économie. Cette règle, ainsi que la quatrième met aussi en avant l'aspect nettement conservateur du SRD. Pour minimiser l'effort cognitif il faut faire appel à des heuristiques que le SRD utilise usuellement. Comme nous allons le voir sous peu, c'est ce qui se passe lors de l'évaluation des alternatives dans le cadre de décisions habituelles. Par ailleurs il est aussi remarquable que le but du SRD dans le cadre de décisions inhabituelles est de les transformer en décisions habituelles via un mécanisme d'apprentissage que nous analyserons bientôt. Il faut partir du principe qu'en contexte décisionnel, le but du SRD est toujours de ramener les situations inconnues et inhabituelles à du connu et de l'habituel, ce comportement est en partie lié au fait que la rationalité des sujets est limitée de façon assez stricte.

La dernière règle est liée à la troisième, elle consiste à *maximiser la facilité avec laquelle le sujet peut justifier ses décisions*. Dans ce cadre-là, le SRD aura tendance à élire l'alternative qui porte avec elle les meilleures justifications. Justifications qui peuvent être de différents ordres et de différents domaines : croyances, connaissances, valeurs, etc. Il sera donc important pour le SRD d'éditer les alternatives avec leurs justifications autant que possible, puisque nous pouvons faire l'hypothèse que des alternatives inexplicables, injustifiables, ou très peu justifiables seront évacuées lors de l'évaluation. En fait cette règle rejoint aussi la première règle, car plus une alternative sera précise et plus elle sera justifiable. La précision peut certes être vue comme le fait pour le SRD d'éditer les alternatives avec leur lot de conséquences, mais aussi pour ce dernier de les éditer avec la (les) meilleure(s) justification(s) possible(s).

+ Évaluation modale et évaluation axiologique.

Le SRD édite les alternatives en termes de possibilités, chaque alternative représente la possibilité qu'un état du monde se réalise plutôt qu'un autre en fonction du but délimité lors de la visée. Dans la phase d'évaluation, un degré de possibilité est attribué à chaque alternative puisque pour une décision donnée, toutes les alternatives n'auront pas le même degré de réalisation possible. Nous appelons ce phénomène *évaluation modale*, puisque la possibilité est une modalité. Ce degré de possibilité est directement connecté aux moyens informatifs et physiques à mettre en œuvre pour la réalisation du but et qui sont représentés par le contenu

sémantique de l'alternative. Moyens qui peuvent être directs ou indirects. Le degré de possibilité reflète la simplicité, la complexité mais aussi l'efficacité et la fiabilité de l'alternative conçue comme un moyen. Possibilité, faisabilité sont par définition susceptibles de degrés, et il convient de les distinguer des probabilités qui peuvent accompagner l'édition des alternatives dans certains cas, relativement peu nombreux. Dans un contexte probabiliste ce sont en général les alternatives avec les probabilités les plus hautes qui vont être choisies. Dans un contexte non probabiliste, l'évaluation de la possibilité d'une perspective se fait en terme de « Chance de », notion qui nous permet de penser la façon approximative dont les sujets se représentent les probabilités et possibilités, c'est d'ailleurs un terme qui est utilisé assez usuellement en contexte décisionnel et qui reflète une estimation plus vague que celui de probabilités.

Le type d'évaluation, en terme de probabilité ou de simple possibilité va aussi dépendre de la catégorie dans laquelle se situe la décision. Pour une décision habituelle ou inhabituelle non critique, une approximation en termes de possibilité suffira. Cependant il faut noter que la plupart des décisions habituelles ne reposent sur aucune évaluation de possibilité particulière, puisque l'habitude permet justement au sujet de « prévoir » et de « projeter » les effets de ses choix usuels sur le monde. Par contre pour des situations critiques ou exceptionnelles le SRD va tenter de construire une mesure probabiliste pour chaque alternative qui soit la plus précise possible. C'est ce qui se passe, par exemple, pour les décisions médicales importantes, ou pour tout choix engageant des éléments centraux de l'existence. *Le degré d'investissement du SRD pour construire ou obtenir une information probabiliste fiable par rapport à une alternative donnée est directement proportionnel à l'enjeu de la décision*, contrairement à ce qui semble acquis en théorie classique de la décision à savoir que les sujets construisent spontanément une mesure probabiliste précise pour tous les types de choix.

Il convient ici de revenir brièvement sur la façon dont le SRD envisage le probable et le possible¹. Nous allons à cet effet nous appuyer sur nos analyses précédentes concernant la pondération des probabilités dans la TP et sur nos analyses du concept de « chance de », à propos duquel nous avons émis l'hypothèse qu'il subit la même pondération qu'une mesure

¹ Pour une analyse de ces concepts voir Granger (1995), p 129-165. L'article de Tversky/Wakker (1995) est lui aussi d'une certaine façon consacré à l'étude de ces notions, ou du moins à la façon dont peut formaliser comment le sujet pondère probabilité et possibilité.

probabiliste classique¹. En fait, les choses sont assez simples, lors de l'évaluation modale, le SRD va se comporter exactement de la façon prédite par la TP : sous-estimation des probabilités/possibilités hautes et surestimation des probabilités/possibilités basses. Ce qui fonctionne pour le domaine des probabilités dans la TP fonctionne ici aussi pour le concept de possibilité représenté par la notion de « chance de ». Ce qu'il faudrait arriver à déterminer totalement c'est quel effet ou influence ce phénomène peut avoir sur l'heuristique d'élimination dont nous parlions plus haut.

La différence du probable avec le possible se situe non pas dans leur pondération, mais dans la teneur informative de chacun de ces concepts. Est possible un événement ou un phénomène qui peut se produire mais dont le sujet ignore précisément à quel point et dans quelles limites, la possibilité ne garantit aucunement la réalité, elle représente une virtualité. Le probable partage avec le possible cette dernière caractéristique avec cette différence qu'il se situe sur une échelle numérique précise, échelle qui donne justement à la possibilité sa mesure et sa limite, et en ce sens nous pouvons dire que le probable mesure le possible, qu'il en est la stricte délimitation. Il y a une différence nette, par exemple, entre : il est possible que ce train soit en retard, et, il est probable que ce train soit en retard. Il s'agit d'une différence de contenu informatif qui est importante dans le cadre de la phase d'évaluation.

Il faut tenter de préciser maintenant à partir de quels éléments le SRD, lors de la phase d'évaluation, attribue-t-il des degrés de probabilité ou de possibilité aux alternatives. Pour les mesures probabilistes, comme nous l'avons vu plus haut, lorsqu'elles sont disponibles elles sont construites lors de la phase d'édition, elles constituent alors une partie de l'information qui constitue l'alternative. La phase d'évaluation les pondérera de la même façon que dans la TP. En ce qui concerne la possibilité, les choses sont un peu différentes, puisque la possibilité des alternatives est estimée lors de la phase d'évaluation. La SRD édite les alternatives comme des possibilités, mais ne leur attribue pas immédiatement un degré de réalisation potentielle, ce travail revient à l'évaluation. La question est : à partir de quels éléments le SRD effectue cette opération ?

Deux éléments vont entrer en ligne de compte ici : l'état présent du SRG et l'état du monde. La possibilité d'une alternative sera déterminée à partir de sa *compatibilité* avec l'état

¹ Voir notre tableau, p 280.

du monde et avec l'état présent du SRG. Pour être minimalement possible, il faut tout d'abord qu'une alternative puisse s'inscrire en *continuité* avec ces deux éléments. Il faudra donc que l'alternative soit pertinente pour la situation donnée. Une alternative est pertinente lorsque par exemple elle permet d'envisager la réalisation du but donné avec des moyens déterminés, mais les moyens à mettre en œuvre doivent être réalistes, et aussi écologiques que possible, plus une alternative sera économique dans les moyens pour un effet maximum et plus son degré de possibilité sera élevé. Il faut garder à l'esprit les analyses de Gigerenzer et Al. Autrement dit, plus une alternative sera coûteuse quant aux ressources physiques et matérielles, et moins elle sera estimée comme possible. Encore une fois il faut noter que ce phénomène dépend aussi du type de décision en question. Dans son fonctionnement habituel, le SRD cherche à être aussi économique que possible, c'est ce qui ressort des quatre règles que nous avons présenté plus haut. Mais ce qui vaut pour l'habituel ne vaut pas forcément pour les situations critiques, inhabituelles ou exceptionnelles, dans lesquelles l'économie quant aux moyens sera moins importante. Ce qui signifie que dans la mesure où nous ramenons le degré de possibilité d'une alternative en contexte habituel à l'économie potentielle qu'elle permet de faire à l'ensemble du système, c'est-à-dire à son aspect écologique au sens de Simon (1955) et Gigerenzer/Brighton (2009), le concept de possibilité ne sera pas le même dans les décisions inhabituelles, critiques ou exceptionnelles.

Le deuxième paramètre par rapport auquel le sujet va estimer la possibilité d'une alternative c'est l'état présent du SRG, c'est-à-dire sa configuration au moment où le choix doit être opéré. Nous appelons « *évaluation axiologique* » le phénomène qui consiste pour le SRD à estimer le degré de compatibilité entre une alternative et la structure générale du SRG. « Axiologique » est ici entendu dans un sens large qui inclut aussi bien l'état affectif du sujet que toutes ses « valeurs », morales ou pas. Comme nous l'avons vu plus haut, le SRG est constitué par l'ensemble des éléments représentatifs, explicites ou non, actuels ou non, que le sujet contient en lui. Ce sont par exemple les croyances, les valeurs, les souvenirs, les connaissances. Tous ses éléments sont liés en un système et la façon dont les éléments sont reliés entre eux détermine l'irréductible singularité du sujet. Le but de l'évaluation axiologique est de mettre en rapport les alternatives et cette singularité sur le mode de la possibilité. La question qui résume le processus est celle-ci : « Cela est-il possible *pour moi* ? », il s'agit de se demander dans quelle mesure l'alternative est compatible avec la configuration du SRG et de

connecter l'irréductibilité de la structure du SRG avec la singularité de la décision. Cela permet d'expliquer par exemple, pourquoi dans deux situations identiques, deux sujets peuvent opter pour des alternatives tout à fait différentes.

Nous incluons dans l'évaluation axiologique le fait, pour le SRD d'évaluer les conséquences potentiellement indésirables des alternatives, notamment celles qui pourraient provoquer des émotions négatives¹. Comme nous l'avons vu plus haut, le SRD cherche à éviter les émotions négatives aussi bien dans son fonctionnement présent que dans l'anticipation du futur. C'est lors de la phase d'évaluation que se fait l'anticipation des conséquences des alternatives, encore une fois, cette anticipation n'est pas systématique, mais sera tributaire du type de décision. Plus le sujet s'éloigne des décisions habituelles non critiques et plus le SRD tentera d'anticiper les conséquences potentielles portées par les alternatives. Comme nous l'avons vu plus haut avec les quatre règles, il s'agit pour le SRD de respecter une certaine « logique » subjective, qui est certes métaphorique, mais qui permet de garder une cohérence sur le court et le long terme. Cette cohérence peut être représentée comme suit : l'alternative choisie doit être compatible avec les valeurs du sujet – nous allons voir immédiatement que cette compatibilité peut avoir des degrés –, et compatible avec l'environnement et l'état du monde, elle doit donc être pertinente.

Dans le cadre de la phase d'évaluation axiologique deux cas problématiques peuvent se rencontrer qui n'ont ni le même impact existentiel, ni le même impact philosophique, il s'agit de deux phénomènes connus : *négozier avec ses valeurs*, et, *agir contre ses valeurs*, contre ce que le sujet juge être le meilleur, c'est-à-dire développer un comportement akratique. La négociation (« trade-off ») du SRD avec les valeurs² du sujet est un phénomène assez courant lorsque nous ne sommes pas dans le domaine de l'éthique ou de la morale³, puisque d'un point de vue général il s'agit pour le SRD de sacrifier certaines préférences subjectives au profit d'autres jugées sur le moment plus importantes. C'est l'essence discriminatoire de la décision

¹ Il s'agit donc de pratiquer une forme d'« affective forecasting », c'est une opération délicate, car les émotions sont des phénomènes difficiles à anticiper, pour de multiples raisons, notamment le fait qu'elles dépendent en partie des préférences et que celles-ci peuvent évoluer avec le temps. En fait les choses sont plus compliquées que cela, le rapport entre émotions et préférences devrait faire l'objet d'une investigation exhaustive, car il est central pour toute philosophie de la volonté.

² Nous rappelons que nous employons le terme valeur dans le sens général défini plus haut qui englobe les valeurs morales et éthiques, mais aussi l'état affectif.

³ Pour l'aspect courant du phénomène on peut se reporter à Bettman/Luce/Payne (1998). Il y a une hiérarchie des valeurs subjectives selon les degrés de négociation possibles. Il y a par exemple ce qui vaut pour le sujet relativement, contextuellement, et ce qui vaut pour lui absolument, inconditionnellement.

et de la phase d'évaluation que d'élire une alternative au détriment des autres qui peuvent elles aussi faire l'objet d'un investissement subjectif. Il est courant que le sujet doive négocier avec ses préférences, ses désirs, ses envies, et même avec ses émotions, bref, avec ce qui définit l'état présent du SRG. Ce qui est paradoxal dans ce type de situation, c'est que le sujet peut avoir l'impression de ne pas faire ce qu'il veut parce que justement il doit sacrifier quelque chose pour être en adéquation avec la situation qui se présente. Or c'est tout à fait faux puisque ce processus repose en fait sur une décision d'ordre deux qui consiste à éditer et à évaluer la négociation en question comme l'alternative la plus pertinente.

Il est important de noter que *le degré de compatibilité d'une alternative avec le système axiologique du sujet est proportionnel au degré de négociation impliquée par l'alternative en question*. Une alternative qui ne fait l'objet d'aucune négociation est parfaitement compatible alors qu'une alternative qui coûte au sujet une négociation trop importante quant à ses valeurs sera évaluée comme peu compatible, voire impossible : son coût semblera exorbitant¹. Le cas s'aggrave lorsque nous entrons dans le domaine des valeurs morales ou éthiques, puisque la négociation avec ce type de valeurs entraîne immédiatement la question de la fidélité à soi-même et aux autres, aux valeurs personnelles et partagées. Non pas que les valeurs morales ne puissent évoluer et se modifier au cours d'une vie, mais elles peuvent difficilement faire l'objet de négociations trop importantes quand elles sont vécues. Elles sont d'ailleurs tellement peu négociables qu'elles peuvent mener au sacrifice² et au don de soi, qui sont la forme ultime du souci de soi. Nous pourrions dire que c'est parce que la perte est jugée trop importante que le sujet peut en arriver à se sacrifier. Le sacrifice est une forme de décision à la limite qu'il faudrait analyser de façon exhaustive.

+ L'évaluation en termes de gains et de pertes et la problème de l'akrasie.

Comme nous l'avons vu dans notre étude de la TP, cette théorie permet d'expliquer un phénomène central pour nos analyses : l'aversion aux pertes. Une des particularités de ce phénomène, mis en avant par KT dans leur article de 1979 – mais aussi bien avant cela par

¹ Voir Payne/Bettman/Johnson (1993). Le coût de l'alternative peut être aussi lié à l'information convoquée pour la construire. Dans certains cas, si la quantité de moyens informatifs et physique à mettre en œuvre pour la réalisation de l'alternative dépasse un certain seuil – qui est toujours singulier – elle sera éliminée lors de la phase d'évaluation comme irréaliste.

² Concept que nous avons déjà croisé et qui pourrait aussi être analysé si nous prolongions nos réflexions sur le travail de la phase d'évaluation dans le domaine de l'éthique et de la morale.

Friedman/Savage (1948) et Allais (1953) – , c’est que, rapportées à l’avoir présent du sujet, les pertes pèsent plus lourd que les gains¹. D’autre part nous avons analysé dans nos parties précédentes l’importance du concept d’« état présent » du SRG et du SRD. Le passage de l’avoir présent à l’état présent s’opère à partir du moment où nous sortons du domaine strictement économique et monétaire pour entrer dans le domaine psychologique et cognitif. Ce passage est rendu nécessaire par le fait que si nous voulons – à partir des résultats de la TP – construire une théorie de la volonté qui soit capable d’intégrer le plus d’éléments représentatifs possibles il faut dépasser le point de vue économique pour rendre justice à tous les éléments cognitifs qui entrent en jeu dans la prise de décision.

Lors de la phase d’évaluation, le SRD va évaluer les alternatives éditées comme des pertes et des gains potentiels par rapport à l’état présent du SRG. Comme nous l’avons vu précédemment la procrastination et dû à une surévaluation de l’état présent du SRD dans le codage des alternatives, les alternatives les plus conservatrices sont conçues comme des moyens de repousser le moment de la décision ou de l’action et sont évaluées comme des gains. L’utilité attribuée à l’état présent du SRG est jugée supérieure à toute modification. Le sujet choisit donc de ne pas agir, ou de ne pas décider « pour le moment », car il aura toujours le temps. Dans notre théorie, nous faisons l’hypothèse que, dans certains cas, le phénomène de l’akrasie peut lui aussi être ramené à une *surpondération de l’état présent du SRG*, mais qui cette fois doit être pensée en relation avec le fait mis en avant par la TP que les pertes pèsent plus lourd que les gains. Les phénomènes de surévaluation de l’état présent et d’aversion aux pertes permettent d’expliquer une certaine forme d’akrasie.

En tant que problème philosophique la notion d’akrasie, c’est-à-dire *d’action intentionnelle contre ce que le sujet juge le meilleur*, a une longue histoire qui commence dans le *Protagoras* et le livre VII de l’*Éthique à Nicomaque*². D’un point de vue classique l’akrasie est causée par l’intervention des passions et des émotions dans le registre de la décision, il s’agit souvent de la décision morale³. Chez Platon l’akrasie est connectée à deux éléments

¹ Voir K/T (1979), p 34. Ce phénomène est représenté mathématiquement par la forme de la fonction de valeur, dont la pente est plus accentuée (« steeper ») pour les pertes que pour les gains.

² 358 C-D ; 1147 a 28. Les réflexions séminales d’Aristote sur l’akrasie sont liées à la notion de syllogisme pratique.

³ Outre les exemples classiques d’Ulysse repris par Elster dans son livre *Ulysses and the syrens*, et de Médée, qui inspirera d’ailleurs Davidson lui-même, puisqu’il en viendra à poser un « principe de Médée », nous trouvons dans la grande littérature classique des exemples « canoniques » de comportements akratiques. C’est le cas par exemple de l’amour-passion de Francesca pour Paolo dans *La divine comédie* de Dante, que Tappolet (2003) analyse finement.

importants : sa thèse morale intellectualiste selon laquelle « Nul n'est méchant volontairement » d'une part, et d'autre part au fait – mis en avant par Protagoras lui-même – que l'on peut « être vaincu par les plaisirs »¹. Aristote de son côté associera l'akrasie à une forme d'impulsivité ou de perte *partielle* de contrôle sur soi, puisque sinon selon le Stagirite, nous passerions du registre de l'akrasia à celui de la mania, c'est-à-dire de la folie, or l'akratès doit être capable d'intentionnalité, c'est essentiel.

Les choses en resteront à peu près là jusqu'à Davidson² qui analysera le phénomène de deux façons différentes : tout d'abord en termes de différence et de désaccord entre deux types de jugements évaluatifs : « toutes choses considérées » et « inconditionnel »³. Le premier type de jugement peut être défini comme suit selon Tappolet (2003) : « It tell us not what we ought to do, but what our reasons indicate we should do »⁴, cela signifie que ce jugement peut ou pas être contraignant quand à l'action, il est prospectif quant aux raisons d'agir.

C'est le premier type de jugement qui va conduire le sujet au comportement akratique, puisque c'est seulement à ce type de jugement que s'applique le principe d'intentionnalité : « One can suppose that such a judgment is the outcome of deliberation »⁵, c'est donc ce type de jugement qui peut motiver l'action akratique, le deuxième type étant en quelque sorte inviolable puisque inconditionnel. Dans une deuxième analyse du problème, Davidson postulera la divisibilité du sujet selon que ses intentions sont contradictoires⁶. La définition davidsonienne de l'akrasie est la suivante :

« In doing *x* an agent acts incontinentely if and only if : (a) the agent does *x* intentionally ; (b) the agent believes there is an alternative action *y* open to him ; and (c) the agent judges that, all things considered, it would be better to do *y* than to do *x* »⁷.

Il s'agit pour le sujet d'agir, et donc de décider de façon intentionnelle et explicite contre ce qu'il considère comme la meilleure option, contre ce qu'il conçoit ou voit comme le

¹ Le *Timée* 85B-87C, associera ce phénomène à une forme de « mania », c'est-à-dire de « folie ».

² Voir Engel (1984), p 14, et Tappolet (2003), in Stroud/Tappolet (Eds.) (2003), p 99.

³ Voir Davidson (1970), p 34-42 ; Tappolet (2003), p 99 ; Engel (2007), p 4-6.

⁴ Tappolet (2003), in Stroud/Tappolet (Eds.) (2003), p 100.

⁵ Ibid.

⁶ Voir Davidson (1986) ; Engel (2007), p 6-8.

⁷ Davidson (1970), p 22. Pour une analyse exhaustive des problèmes reliés à l'akrasie, ainsi qu'aux différents types d'akrasie on peut se reporter au livre édité par Stroud/Tappolet (2003), ainsi qu'aux articles de Mele (2009) et May/Holton (2010).

meilleur (« video meliora... »), ainsi que l'exprime la phrase d'Ovide : « je vois les choses les meilleures, et je les éprouve comme telles, mais je fais les pires »¹ (« Video meliora proboque deteriora sequor »). L'akrasie a longtemps été considérée comme un phénomène irrationnel puisque le sujet agit contre ses propres intérêts et contre ce qu'il *croit* être le meilleur².

Nous allons proposer ici une explication de l'akrasie originale, puisque jamais tentée à notre connaissance³, fondée sur trois phénomènes issus de la TP que nous reconstruisons philosophiquement : aversion aux pertes, les pertes pèsent plus lourd que les gains, surévaluation de l'état présent. Dans notre théorie, tous ces phénomènes sont à ramener à la phase d'évaluation. Nous pensons pouvoir expliquer à l'aide de ces concepts retravaillés un type d'akrasie liée avec l'aspect dynamique du SRD, il s'agit de l'impossibilité de passer d'un état A à un état B, jugé meilleur à tous les points de vue. Ce type d'akrasie est conditionné par la préférence des sujets pour le statu quo qui est une conséquence des analyses que nous avons menées dans la lignée de la TP. Ce phénomène doit être mis en relation avec le concept d'inertie que nous avons croisé à plusieurs reprises, et qui est lié à l'aspect statique du SRD. Comme nous le remarquons il y a une ambivalence dans l'aspect conservateur du SRD, car il permet aussi bien au sujet d'être cohérent dans ses décisions habituelles, qu'il peut être la cause d'un type de comportement akratique à savoir l'inertie décisionnelle pathologique. Le SRD doit trouver un équilibre entre la cohérence et l'inertie qui ajoutée à l'aversion au risque, peut causer une résistance pathologique au changement.

La question à laquelle il faut tenter de répondre est la suivante : pourquoi un sujet préfère opter pour un comportement X ou sa continuation, alors qu'il sait que si il changeait pour un comportement Y son état général s'améliorerait. Pourquoi par exemple le sujet continue de fumer alors qu'il sait pertinemment que l'arrêt du tabac serait meilleur pour lui ? Ici, comme dans le cas de Davidson nous avons affaire à deux jugements évaluatifs : « je fais X mais je sais que c'est mauvais », « je change pour Y et je sais que c'est bon ». Seulement le ressort du comportement akratique va se jouer ailleurs que dans le désaccord entre deux types

¹ Ovide, *Métamorphoses*, 7 : 20.

² Il existe bon nombre d'interprétations, internalistes ou externalistes du phénomène, dans lesquelles nous ne rentrerons pas. Il existe aussi des auteurs comme Holton (1999) qui nient que l'akrasie soit un échec de la volonté, d'autres comme Hare (1952) et (1963), dénie que le phénomène soit en tant que tel possible, puisque selon lui tous nos jugements évaluatifs sont prescriptifs : nous agissons donc toujours en fonction de nos évaluations.

³ Ce qui ne veut pas dire qu'aucune tentative n'a été faite pour construire une explication de l'akrasie à partir de la théorie de la décision, ce dont témoigne par exemple Jackson (1984), et Jeffrey (1983), p 214-224.

de jugement évaluatif, parce que dans notre exemple les deux jugements ont la même pertinence quant à l'action, ils sont tous les deux aussi contraignants, mais pour des raisons différentes. Dans le cas qui nous occupe, c'est *l'aversion aux pertes induite par la représentation du changement*¹ qui va jouer un rôle déterminant. Comme nous l'avons vu plus haut, c'est un phénomène très robuste et qui a pour conséquence immédiate que toute perte est surévaluée par rapport à un gain identique, voire même légèrement supérieur, et cette surévaluation peut être très importante. Dans le cadre de notre fumeur qui sait que l'arrêt du tabac est meilleur pour lui, le SRD éditée cette alternative comme un gain, continuer de fumer est éditée de façon mixte comme une perte quant à la santé, mais un gain quant au plaisir pris au comportement². Du fait que la perte est systématiquement surévaluée³ elle peut entraîner de pair une surévaluation des conséquences négatives de l'arrêt du tabac (prise de poids, irritabilité, perte sèche du plaisir de fumer, etc.), et causer ainsi le statu quo⁴. Nous retrouvons ici aussi, comme dans la procrastination une surévaluation de l'état présent du SRG par rapport à son état futur potentiel : l'état présent est éditée comme un gain dont les conséquences néfastes, bien que connues sont diluées dans un futur qui demeure de toute façon opaque au décider. À partir de ces considérations nous pouvons dire que pour ce type d'akrasie ce sera l'alternative éditée comme pouvant garantir le statu quo qui sera élue lors de la phase d'évaluation.

Ce qui ressort de ce type de comportement que nous qualifions d'akratique puisque le sujet décide contre son meilleur jugement, et qui n'avait pas été à notre sens suffisamment mis en avant, c'est l'importance et le poids du phénomène de l'aversion aux pertes⁵ dans l'existence. Ce phénomène permet dans un premier temps d'expliquer l'aspect conservateur du SRD. Il convient donc de lui attribuer la place qui lui revient dans la prise de décision : elle

¹ Ici il conviendrait de travailler dans le détail le rapport du SRD au changement.

² Pour être tout à fait exhaustif, il faudrait considérer les préférences de deuxième ordre. Parce que seules les préférences du premier ordre du sujet sont akratiques ici. Il est assez raisonnable de faire l'hypothèse que le fumeur préfère ne pas préférer fumer à préférer fumer, mais justement dans un contexte de deuxième ordre comme nous situons à un niveau plus élevé de préférences la contrainte sur le comportement est moins forte. Il serait toutefois important d'analyser à fond et de comprendre comment les préférences et croyances de second ordre peuvent aider à comprendre certains types d'akrasies. On trouve un excellent exemple de ce genre d'explication dans Jeffrey (1983), p 214-224. L'exemple est justement celui du fumeur.

³ La perte sera toujours plus fortement surévaluée selon que le comportement qui doit être abandonné sera pourvoyeur de « plaisir ».

⁴ Il est remarquable que dans ce contexte-là, les conséquences négatives du « bon comportement » soient elles-mêmes surévaluées par rapport aux conséquences positives, à ce point qu'elles peuvent même l'emporter dans la décision. Il y a beaucoup de sujets qui préfèrent ne pas s'arrêter de fumer, plutôt que de prendre du poids.

⁵ Cependant le phénomène très répandu de l'achat d'assurance, parfois à n'importe quel prix et pour n'importe quel type d'objet aurait pu nous mettre sur la voie.

permet par la promotion implicite qu'elle fait de l'état présent du SRG de stabiliser le SRD autour de certaines habitudes, mais c'est une stabilisation qui n'est pas toujours garante de la plus grande cohérence puisqu'elle peut aussi mener à des comportements akratiques. Il est crucial pour le système de référence décisionnel de trouver un équilibre entre la stabilité et l'inertie, parce que certaines situations exigent que les sujets puissent s'adapter et sortir de leurs habitudes et références habituelles, que ce soit de façon progressive ou de façon rapide.

b – Les différents types d'évaluation et le concept d'apprentissage.

+ L'évaluation automatique.

Il existe un type de décision qui se situe à la limite de la conscience et de la cognition, à la limite de la représentation. Il s'agit des décisions que nous avons désignées dans notre tableau sous le terme presque oxymorique d'« automatismes », oxymorique puisque le concept de décision est généralement associé à un ensemble de processus. Peut-être aurait-il été plus pertinent de parler de décisions instantanées. Il existe donc des décisions qui ne sont pas à strictement parler processuelles, et dont les phases d'établissement de la visée et d'édition sont difficilement discernables. Dans ce cadre-là, l'évaluation va essentiellement consister à actualiser ou mobiliser les éléments habituels du SRG pour répondre à la situation. Il s'agit ici des dizaines de *microdécisions* que nous prenons chaque jour sans y faire attention, des décisions qui se situent à la limite du réflexe, mais qui ne peuvent être exclues purement et simplement de la sphère du décidable. À partir du moment où nous ne pouvons plus distinguer de visée et d'édition pourquoi pouvons-nous dire que nous sommes encore dans le registre du décidable ?

Nous nous situons ici encore dans le registre du décidable car bien que le sujet réponde automatiquement à la situation d'une façon donnée en optant pour une manière de faire habituelle ou une manière d'être, il existe toujours d'autres alternatives possibles que le sujet pourrait choisir d'actualiser. Par exemple imaginons un individu qui se ronge les ongles de façon compulsive en contexte de stress¹, c'est sa réponse instantanée à ce type de situation, il se ronge les ongles sans établir de visée ni éditer sa pratique, cependant elle va être évaluée de

¹ Pour une analyse intéressante de ce type de comportements situés aux limites du décidable, voir Ainslie (2005).

façon presque inconsciente comme une forme de réconfort immédiat. Peut-on dire pour autant qu'elle n'est pas le fruit d'une décision ? Non, car le sujet peut par exemple choisir de refréner son comportement habituel. Ce qu'il faut retenir ici c'est que nous sommes dans le domaine du décidable dès le moment où il existe plusieurs alternatives. Or dans la plupart des situations où les décisions se font automatiquement, il existe d'autres alternatives possibles pour lesquelles le sujet peut opter. Le feu devient orange, le sujet ralentit de façon automatique, mais il pourrait aussi bien accélérer pour tenter de passer quand même. Nous associons les décisions instantanées au registre des manières d'être ou de faire car elles constituent la réponse habituelle à des situations qui ne demandent que très peu voire pas du tout d'élaboration de la part du SRD. Ces situations nous permettent aussi d'évaluer comment le SRD appuie son fonctionnement, même lorsque celui-ci est réduit au minimum, sur les habitudes. Car nous pourrions dire qu'une façon de faire ou une façon d'être c'est la manifestation de l'utilisation des habitudes par le SRD, et cette utilisation est tellement coutumière et usuelle que souvent le sujet ne s'en rend même pas compte, et qu'il serait probablement difficile, la plupart du temps, pour lui de se rappeler quand il a acquis ce type de comportement. Penser ces comportements, ces manières d'être ou de faire, en termes de décisions permet d'élargir ce concept et d'y intégrer des notions comme l'automatisme ou l'instantanéité qui peuvent de prime abord sembler totalement antinomiques avec le concept de délibération par exemple. Il existe un ensemble de décisions non délibératives et à la limite de la conscience qui constituent la limite inférieure du décider quant aux moyens et à l'investissement du SRD. Bien entendu nous nous situons ici dans le registre de l'habituel et du coutumier, et l'investissement du SRD quant aux fins et aux moyens sera d'autant plus important que nous sortons de ces registres.

+ L'évaluation heuristique : l'application du principe d'économie.

Il est temps de passer à l'analyse de ce que nous pourrions appeler « les heuristiques d'évaluation ». En contexte habituel, lorsque nous sortons des décisions automatiques, le SRD va évoluer et osciller autour de certaines habitudes et de certaines heuristiques qui constituent son mode de fonctionnement usuel et assurent la cohérence du sujet quant aux décisions, sur le court et sur le long terme. Pour arriver à cerner le phénomène complètement il faut revenir brièvement sur la théorie des heuristiques et des biais de Kanheman et Tversky ainsi que sur la

théorie de Gigerenzer. D'un côté avec la théorie de K/T nous avons un pessimisme affiché quant à la rationalité des sujets : l'usage des heuristiques est conçu comme une défaillance par rapport à certains critères de rationalité. De l'autre côté, avec sa théorie de la rationalité limitée et sa théorie des heuristiques Gigerenzer propose une vision certes plus réaliste de l'usage des heuristiques, mais peut-être un peu trop optimiste selon nous. Il faut donc parvenir à nous frayer une voie philosophique entre le pessimisme affiché de KT et l'optimisme de Gigerenzer.

Dans notre théorie, heuristiques et habitudes sont liées en ce sens que les heuristiques que va mettre en place le sujet au fur et à mesure de son existence pour ses décisions quotidiennes vont se changer en habitudes du fait que le sujet n'est pas toujours en train de rechercher ou d'élaborer des méthodes de décisions, cette recherche s'inscrit dans le cadre des décisions inhabituelles, critiques et exceptionnelles. Ce phénomène est lié au fait que les situations habituelles qui se présentent aux sujets peuvent se répéter, et que cette répétition les induit à réutiliser les mêmes heuristiques qui deviennent ainsi des habitudes¹. Dans ce contexte-là il est important de noter aussi que les habitudes décisionnelles iront en se consolidant, voire en se sclérosant au fur et à mesure que le sujet va prendre de l'âge.

Dans le cadre des décisions habituelles, nous allons limiter le fonctionnement du SRD à une heuristique principale qui peut produire des dérivées. L'idée sur laquelle repose cette heuristique permet de trouver une voie moyenne entre la théorie de K/T et celle de Gigerenzer, et même de les subsumer, quoique dans son principe elle soit plus proche de la théorie de Gigerenzer. Il s'agit de ce que nous pourrions appeler une « heuristique d'évaluation » de second ordre : lors de la phase d'évaluation des alternatives, le SRD utilise et mobilise toujours les heuristiques qui permettent le plus de confort intellectuel, le plus cohérence, et surtout qui nécessitent le moins d'efforts pour obtenir le plus de moyens. Comme nous le voyons cette heuristique est en adéquation avec les quatre règles que nous avons présentées plus haut et avec le fait que dans son fonctionnement habituel, le SRD cherche à être le plus conservateur possible.

¹ D'où la maxime d'Aristote, issue de l'*Éthique à Nicomaque* : « L'habitude est une seconde nature ». Ce qui ne veut pas dire que le SRD ne peut pas modifier son comportement heuristique pendant la phase d'évaluation en contexte habituel, mais ce sera seulement dans la mesure où la modification envisagée est un gain net quant aux fins et aux moyens, n'oublions pas que le SRD fonctionne selon le principe d'économie et que plus une alternative sera économique et plus elle aura de chance d'être réalisée, il en ira de même avec les heuristiques : si une heuristique plus économique semble pouvoir être envisagée elle sera automatiquement sélectionnée par une « heuristique d'évaluation ».

L'aspect économique et écologique est donc capital lors de la phase d'évaluation mais il faut préciser un peu les choses. Nous avons vu plus haut que les alternatives étaient évaluées quant aux moyens qu'elles impliquent pour leur réalisation et que celles qui nécessitent des moyens trop élevés étaient la plupart du temps écartées ou éliminées dans un contexte habituel. Il faut ici un peu approfondir le concept d'évaluation écologique et économique. En contexte habituel, la principale heuristique dérivée de celle que nous avons vu plus haut est très proche de la « recognition » de Gigerenzer¹, mais elle se rapporte à des éléments plus précis du SRG : elle consiste à évaluer les alternatives en fonction de la mémoire affective et de la croyance en la régularité du quotidien. Nous avons vu que dans l'édition des alternatives, c'était l'affectif qui venait en premier, il en va de même lors de la phase d'évaluation. Dans le cadre de décisions habituelles c'est la confrontation des alternatives avec la mémoire qui va être déterminante, mais la mémoire qui va être convoquée en premier lieu sera la mémoire affective se rapportant à la situation. Il s'agit pour le SRD de maintenir une cohérence et une continuité à tous les niveaux, et en particulier dans la dimension affective. Par exemple si le sujet est dans un état neutre, le but du SRD dans la phase d'évaluation sera de maintenir cet état et ne pas sélectionner des alternatives qui pourraient venir altérer cet état dans un sens négatif : seront favorisées les alternatives qui peuvent maintenir cette neutralité ou faire passer le sujet dans un affect positif.

Le travail heuristique dans la phase d'évaluation en contexte habituel consistera essentiellement à mettre en rapport les alternatives avec la mémoire du sujet, ce phénomène permet aussi de mieux comprendre comment sont évaluées les alternatives quand à leur faisabilité et leurs possibilités. Une alternative sera jugée d'autant plus possible et réalisable qu'elle sera *reconnue*² comme pertinente lors de situations passées identiques ou quasi identiques. Le SRD pratiquera donc une forme d'induction à partir d'indices, comme dans la théorie de Gigerenzer. La différence étant que ce n'est pas toujours le même type de mémoire qui est convoquée, et que les différents types de mémoires sont hiérarchisés de la même façon que le sont dans la phase d'édition les différents éléments cognitifs : à partir du codage affectif et du codage doxique. La question qui se pose c'est dans quelle mesure ce type de

¹ Voir Goldstein/Gigerenzer (2002), in Gigerenzer (Ed.) (2011), p 60-81.

² C'est ici que notre définition de la recognition s'écarte légèrement de celle de Gigerenzer, bien que ici aussi la recognition soit occasionnée par les informations extérieures.

comportement de mène-t-il pas à une forme de paresse intellectuelle qui peut s'avérer handicapante dans certaines situations, comme nous l'avons vu avec l'akrasie, il existe des situations où l'inertie décisionnelle est telle qu'elle provoque une incapacité pour le sujet à changer de comportement, même lorsque celui-ci est toxique. Il convient donc de parvenir à trouver une juste mesure entre l'aspect adaptatif de la phase d'évaluation qui cherche à minimiser l'utilisation des ressources internes ou externes, et les problèmes qui peuvent survenir lorsque le SRD verrouille complètement son mode de fonctionnement. C'est un équilibre difficile à trouver et dont le point est toujours singulier, mais c'est aussi un équilibre nécessaire, car du fait que le SRD est connecté et déterminé par l'état présent du SRG, et que celui-ci évolue, il faut que le SRD soit capable de suivre cette évolution tout au long de l'existence du sujet.

Nos vues permettent aussi d'expliquer le pessimisme de K/T dans leur théorie des heuristiques et des biais : c'est parce que les sujets sont confrontés à des décisions de types habituels dans des contextes quotidiens permettant une forme de paresse intellectuelle qu'ils n'approfondissent pas les données des problèmes qui leurs sont soumis, c'est aussi ce qui se passe dans la TP. Nous pouvons faire l'hypothèse que confrontés à des situations critiques ou ultimes, engageant des enjeux très élevés qui dépendraient de l'évaluation des sujets, ces derniers auraient tendance à analyser de beaucoup plus près les données et d'être plus circonspect, et par définition d'utiliser des heuristiques plus fiables qui produiraient moins de biais. *C'est le quotidien et l'habituel qui conditionnent les sujets à produire des estimations biaisées, puisque le quotidien et l'habituel ne nécessitent guère plus que des estimations et des approximations.* La précision et les évaluations fines sont à relier à l'inhabituel, au critique ou à l'exceptionnel. Notons enfin que le SRD cherche toujours aussi à être économe quant au temps investi.

+ Le SRD et le concept d'apprentissage.

Le système de référence décisionnel désigne un ensemble de processus qui permettent au sujet d'élaborer ses décisions de diverses façons et dans divers contextes. Nous nous sommes ici surtout concentrés sur les décisions habituelles parce qu'elles sont les plus courantes. Il est temps maintenant de nous intéresser à l'aspect dynamique du SRD qui est le complément de son côté statique. Nous entendons ici dynamique dans le sens de capable

d'évoluer par une forme d'apprentissage. Cet apprentissage du SRD est rendu nécessaire, car il existe un certain nombre de situations auxquelles il doit s'adapter et dans lesquelles il ne peut utiliser ses stratégies habituelles.

Il existe deux types de cohérences importants pour nous en contexte décisionnel : sur le court terme, et sur le long terme. C'est surtout sur d'assez longues périodes de l'existence que les stratégies décisionnelles peuvent être amenées à se modifier en profondeur avec l'évolution du SRG. Mais il peut aussi se présenter sur le court terme toute une gamme de décisions qui peuvent obliger le SRD à sortir de son comportement usuel. Habituellement, l'existence repose sur une forme de régularité qui alterne avec des changements plus ou moins importants, changements qui pourraient d'ailleurs être hiérarchisés selon leur impact. Si le SRD fonctionne normalement il doit pouvoir faire face à toutes les situations de l'existence aussi bien à son aspect régulier et habituel qu'aux divers changements qui peuvent advenir, et qui peuvent aller de très petits changements dans le registre du quotidien, à des changements plus importants, voire ultimes, engageant l'ensemble de l'existence ou compromettant son intégrité. L'établissement d'une éventuelle échelle ou hiérarchisation des changements – provoqués ou subis – devrait être connectée à la dynamique du système de référence général, car c'est elle qui reflète l'intégration du changement existentiel dans la subjectivité, et qui a un impact immédiat sur le fonctionnement du SRD.

Lorsqu'une situation inhabituelle se présente, la première opération que le SRD va effectuer lors de la phase d'évaluation sera de convoquer la stratégie habituelle qui semble la plus se rapprocher de la situation, avant de passer à l'opération de production d'une nouvelle stratégie, le SRD va essayer d'appliquer autant que faire se peut les options qui sont déjà disponibles. L'application du principe d'économie sera donc prolongée jusque dans le domaine des décisions inhabituelles. Cependant cette opération peut fonctionner dans certains cas, où l'inhabituel ne se détache que très peu de l'habituel, mais elle peut aussi échouer, poussant ainsi le SRD à élaborer une nouvelle stratégie pour faire face à l'inédit de la situation¹. Lorsqu'une nouvelle stratégie est construite dans la phase d'évaluation, celle-ci sera intégrée immédiatement à l'ensemble des stratégies que le SRD a en réserve afin de pouvoir faire face à la situation si elle se présente de nouveau. Le but du SRD est simple : ramener

¹ Comme nous l'avons vu plus haut il est de la plus grande importance que le SRD soit capable d'auto-régulation et d'une forme de réflexivité, il a donc certaines caractéristiques métacognitives, voir Proust (2007).

l'inédit à des stratégies connues par souci d'économie, sinon produire une stratégie qui permettra à l'ensemble du système de répondre le plus économiquement possible lorsque la situation se reproduira. Auquel cas le SRD pourra l'éditer dans la catégorie de l'habituel non-critique. C'est une première forme de l'apprentissage du SRD qui est à relier à son aspect adaptatif : le SRD doit être capable de se mettre au niveau de la situation, mais comme nous l'avons vu certains comportements « pathologiques » peuvent maximiser la résistance au changement et bloquer le processus¹.

Le deuxième type d'apprentissage du SRD est un peu différent, il est lié à l'aspect critique de la situation. Il y a plusieurs façons d'envisager une situation critique, par exemple selon l'enjeu de la décision ou selon les contraintes qui pèsent sur l'édition et sur l'évaluation, ce peut être par exemple une contrainte de temps². C'est surtout ce dernier exemple qui va nous intéresser en sachant que dans ce contexte aussi les choses doivent être nuancées : il y a des degrés dans les contraintes de temps, du délai long à l'urgence absolue. Le SRD est un système qui fonctionne sur un mode temporel que nous pourrions dire normal, lorsqu'aucun délai n'est présent. Pour faire face à une situation critique quand au délai le SRD doit être éduqué afin de pouvoir travailler sous contrainte de temps. Les sujets sont outillés pour décider dans l'urgence mais ils ne le font pas de façon spontanée, et ceci peut être expliqué par le fait que décider en contexte de délai peut être source de stress, et que ce stress peut être vécu comme une émotion négative. Or nous avons vu qu'un des buts principaux de la phase d'évaluation est d'éviter ce type d'émotions. Par ailleurs nous avons vu que lorsque le délai est trop sévère cela peut amener le sujet à négliger des éléments importants lors de la phase d'édition, comme de l'information, l'exhaustivité des alternatives et l'édition de toutes leurs conséquences.

Pour faire face à des situations de ce genre, qui sont nombreuses dans l'existence, surtout dans le domaine professionnel, il faut éduquer le SRD à fonctionner sous contrainte de temps. Un exemple assez simple de ce type d'apprentissage est la réalisation d'examens en temps limités lors de la scolarité. Dans la plupart des domaines professionnels la contrainte de temps est présente selon des degrés divers : il y a un écart presque incommensurable entre le

¹ Il faudrait aller beaucoup plus loin dans l'exploration de ce phénomène.

² Nous choisissons de faire ici de tout délai un facteur critique puisque qu'il soit long ou pas il ajoute une contrainte sur la décision.

travail à remettre pour une date X et la prise en charge du patient qui arrive aux urgences dans un état critique¹. Pourtant dans ces deux cas qui sont très éloignés l'un de l'autre ce qui permet d'expliquer la possibilité d'une réponse pertinente à la situation c'est la façon dont le SRD a été éduqué lors de l'apprentissage de la profession en question.

La capacité d'apprentissage du SRD, qui comme nous l'avons vu se décline dans plusieurs domaines, devrait faire l'objet d'une investigation à part entière, que nous avons juste eu le temps d'ébaucher ici. Ce qui ressort surtout c'est qu'il y a deux façons pour le SRD d'être en situation d'apprentissage : endogène ou exogène. Face à des situations inhabituelles le SRD se mettra en général à jour de façon autonome, face à des situations critiques quant au délai nous faisons l'hypothèse que l'éducation du SRD se fera de façon plus exogène, grâce à l'entraînement et à l'exercice, en sachant que de toute façon le but ultime du système c'est de ramener l'inhabituel à de l'habituel, le critique au non critique, bref en quelque sorte d'assimiler – au sens aristotélicien défini plus haut – les situations qui viendront nécessairement compromettre sa stabilité tout au long de l'existence.

Du point de vue philosophique il était ici question de rentrer le plus en avant possible dans le fonctionnement de la *volonté conçue comme un ensemble de processus coordonnés dans l'élément de la représentation*. L'originalité de la démarche réside ici en ce que les éléments centraux du système trouvent leur origine dans la Théorie de la Perspective. Nous avons tenté de trouver dans la connexion des trois opérations que sont la visée, l'édition et l'évaluation, avec les éléments représentationnels contenus dans le SRG, « le grain fin » dont Pascal Engel² pensait qu'il était nécessaire pour arriver à cerner d'aussi près que possible la spécificité et la singularité du décidable. Pour ce faire il aura fallu reconstruire l'intégralité des phases d'édition et d'évaluation de façon conceptuelle.

Ce qu'il est important de retenir ici c'est la façon dont fonctionnent l'édition et l'évaluation pour produire une décision : l'édition donne aux alternatives une forme et un

¹ Voir en particulier le très bon exemple développé par Gigerenzer/Todd (1999), in Gigerenzer/Todd and ABC Research Group (Eds.), p 3-5.

² Voir Engel (1997), p 339.

contenu, et l'évaluation rapporte ce contenu à l'état présent du système de référence général et à l'état du monde, afin d'élire l'alternative qui sera d'une part la plus économique quant aux moyens et la plus en adéquation et en continuité avec ces deux éléments. De la même façon que les émotions et les croyances interviennent dans l'édition, elles interviennent aussi dans la phase d'évaluation, mais pas exactement sur le même mode, c'est un point qu'il faudrait approfondir.

Il faut aussi garder à l'esprit l'importance de l'état présent du SRG, qui détermine la totalité du travail du SRD. C'est par rapport à cet état que les alternatives vont être codées comme des gains ou des pertes. C'est un phénomène qui permet de faire ressortir – en partie grâce au concept d'aversion aux pertes et de ses conséquences – la centralité du présent pour l'existence, mais une centralité qui peut aussi se transformer en handicap par exemple dans les phénomènes de procrastination et d'akrasie qui sont tous deux liés à une surévaluation du présent.

Conclusions générales

Le présent travail avait pour but de reconstruire le concept de volonté à partir de la Théorie de la Perspective grâce à l'introduction du concept de Système de Référence Décisionnel – qui est la *généralisation philosophique naturelle* de la fonction de valeur – et dans la relation de ce dernier au Système de Référence Général. Cette reconstruction n'a été possible que parce que ces deux systèmes sont le lieu d'un certain nombre et d'un certain type de représentations cognitives qui ont diverses influences dans la prise de décision. La Théorie de l'Utilité Attendue peut être considérée comme une théorie *minimalement représentationnelle* – en ce sens que les seules représentations qu'elle prend en compte sont les valeurs des issues pondérées par les probabilités –, c'est pourquoi son domaine d'application, bien que réel, ne peut être que très limité et délimité. De son côté la TP arrive à incrémenter dans le formalisme de la TUA l'influence de la pondération des probabilités et de l'avoir présent du sujet, mais comme elle ne sort pas du formalisme ni du contexte économique, elle doit être aussi dépassée vers une théorie qui donne à l'ensemble des éléments représentationnels et cognitifs intervenants dans une décision la place qui leur revient dans le processus du choix rationnel : une théorie conceptuelle et philosophique. Il ne s'agit pas de rejeter en bloc les acquis de la TUA ou de la TP, mais de montrer qu'elles ne peuvent prétendre à expliquer l'ensemble du décidable, à partir du moment où il est devenu clair que la définition du décidable ne peut faire l'économie de la *subjectivité* et d'un concept de *représentation* qui englobe tous les éléments du SRG, en tout cas dans la sphère des décisions humaines.

Des analyses poussées sur les phases d'édition et d'évaluation devraient encore plus détailler comment les alternatives sont construites et ce qui fait qu'elles sont singulières à deux points de vue : quant à la décision et quant à la structure irréductiblement individuelle du SRG. Ces analyses devraient aussi étudier de plus près les mécanismes de l'évaluation, par exemple en s'interrogeant sur la possibilité pour le SRD d'évaluer les alternatives de façon conjointe ou pas¹, et en approfondissant le rôle des heuristiques d'évaluation. Il faudrait aussi

¹ C'est une question importante, car il s'agit en fait de se demander quelle influence les représentations des alternatives peuvent avoir les unes sur les autres. Le problème a été abordé dans l'économie et nous pensons qu'il est transposable dans notre structure théorique de façon pertinente. Il ne faut pas oublier que les alternatives sont souvent éditées comme des ensembles d'options disjointes. Nous pensons que ces éléments peuvent avoir une influence les uns sur les autres.

analyser de façon exhaustive le fonctionnement du SRD dans les situations inhabituelles, critiques et exceptionnelles, afin d'avoir une vue complète de son fonctionnement. Il conviendrait de se demander notamment jusqu'où va l'utilisation des heuristiques, et jusqu'où va par exemple la portée explicative du comportement heuristique des sujets, parce qu'il est clair que plus la situation devient critique et moins le comportement va être heuristique puisque certaines situations ne pourront plus être gérées par des méthodes approximatives.

À partir du travail du SRD dans les opérations de visée, d'édition et d'évaluation, nous avons exploré les éléments centraux de la *volonté pensée sur le mode d'un ensemble de processus coordonnés dans l'élément de la représentation*. Car toutes les opérations que nous avons analysées ici – même l'évaluation automatique – se situent dans l'élément de la *représentation*, qui comme nous l'avons vu plus haut est une forme de *pondération originale* qui trouve sa raison dans la configuration singulière du SRG. Nous pourrions donc dire que le système de référence décisionnel, c'est l'actualisation de la volonté.

Ce travail comporte une partie fondationnelle en ce qui concerne les décisions habituelles non critiques et aussi une partie programmatique puisque toutes les pistes et toutes les problématiques n'ont pas été épuisées. Il conviendra dans les recherches à venir, par exemple de s'interroger, sur l'articulation des principaux éléments du SRD entre eux¹, d'arriver en quelque sorte à expliciter, à l'aide des outils de la logique, de la phénoménologie, mais aussi des données de la psychologie cognitive et de la philosophie de l'esprit, l'ensemble des éléments représentationnels qui sont à l'œuvre selon le type de situation dans lequel le sujet se trouve. Il conviendrait aussi d'arriver à définir et à déterminer toutes les heuristiques d'édition et d'évaluation, et de sonder plus exactement et d'une façon plus approfondie les affinités théoriques entre nos hypothèses et les travaux sur la rationalité écologique inaugurés par Simon (1955) et Gigerenzer (1996) d'une part, il serait aussi important d'arriver à évaluer l'aspect adaptatif du SRD, dans la lignée des travaux de Tooby/Cosmides.

Cette conclusion va s'axer autour de deux questions qui s'inscrivent en continuité avec nos analyses précédentes, mais en ouvrant sur de nouveaux domaines d'investigations qui semblent prometteurs. Tout d'abord il s'agit de se demander où se situe la liberté dans notre

¹ L'hypothèse étant que tous les éléments du tableau p 219 qui font partie du SRG peuvent être développées comme nous l'avons fait pour le SRD. Nous pourrions arriver par là à envisager un véritable Système Général de la Représentation, dont la structure reste à définir.

théorie ? Dans un second temps il va s'agir de se questionner sur les leçons à tirer de cette surpondération ou surévaluation du présent, qui fait du SRD un système cohérent quant au cours habituel de ses décisions, mais qui fait aussi courir au sujet le risque de l'inertie pathologique. Cette interrogation nous permettra de prolonger notre réflexion sur la possibilité d'attribuer à des structures plus importantes, par exemple à l'État, des comportements akratiques ou procrastinateurs.

Il est difficile de penser la volonté sans au moins évoquer le problème de *la liberté*. D'une certaine façon il va être question ici de se demander comment s'opère le passage à une éthique minimale, à savoir une éthique de la liberté, dans le cadre de nos hypothèses. C'est un passage nécessaire pour la raison que nos décisions, bien que la plupart du temps privées¹, se situent dans un espace familial, social, professionnel un espace duquel l'altérité ne peut ni de doit être négligée, et ce sous n'importe quelle forme qu'elle puisse se donner. C'est ce que nous avons d'ailleurs vu en analysant comment autrui peut créer une résistance dans notre espace décisionnel.

La question à laquelle nous allons tenter de répondre ici est la suivante : où se situe la liberté dans le SRD ? Précisons notre question : quels sont les éléments intervenant dans le décider sur lesquels le sujet peut avoir prise afin de donner à la décision le cours qu'il juge pertinent ? En ce qui concerne l'environnement extérieur, le décideur ne peut avoir qu'une maîtrise partielle de ce dernier et des informations qui circulent en lui. Cette maîtrise concerne par exemple l'ordonnancement des choses et des objets et leur manipulabilité, et ce même dans des environnements assez quotidiens, comme nous l'avons vu cette maîtrise peut en effet entrer en ligne de compte dans divers types de décisions, mais à notre sens ce n'est pas là que se situe l'essence de la liberté. Ce qui va dépendre du sujet d'une façon plus nette va se situer à l'intérieur du SRG, nous retrouvons ici la distinction stoïcienne classique entre ce qui dépend de nous et ce qui n'en dépend pas². L'extériorité, l'environnement ne dépendent que très partiellement de nous, voire pas du tout, alors que certaines de nos représentations en dépendent peut-être un peu plus.

À notre sens c'est donc dans le SRG qu'il sera plus pertinent de situer la liberté, plus précisément, nous pensons que *la liberté se situe dans le rapport, ou les rapports, que le sujet*

¹ Nous avons vu comment le contexte hiérarchique pouvait faire exception à cette règle.

² Voir Long/Sedley (2001), p 475-501.

entretient avec ses représentations. Par exemple dans la préséance que le sujet va accorder à certaines d'entre elles, comme les croyances ou les valeurs. D'une certaine façon nous pouvons dire que l'engagement du sujet vis-à-vis d'une croyance particulière ou son degré d'investissement dans une valeur dépendent de lui, et que c'est dans ce degré que se situe sa liberté. Le sujet peut aussi être amené à amender, à modifier certaines de ses croyances ou valeurs, il n'y a pas de nécessité autre que subjective de l'investissement du décideur dans ses croyances ou ses valeurs, à moins de nous situer dans le cadre de l'endoctrinement et du fanatisme. D'un point de vue plus général *la liberté c'est la capacité d'utiliser ses représentations de façon pertinente pour la décision donnée.*

Nous trouvons ici un second point de contact avec le stoïcisme parce que nous situons la liberté subjective dans la capacité de donner son *assentiment* à un certain type de représentations. Il y a donc un rapport entre la liberté et le jugement dans le sens où le sujet évalue telle représentation cognitive comme pertinente pour parvenir à telle fin¹. Cependant il ne faudrait pas non plus accorder au sujet trop de maîtrise sur les représentations qui constituent le SRG, il y a ici aussi des degrés, et il faudrait opérer encore des distinctions entre les représentations cognitives qui sont susceptibles d'être maîtrisées et celles qui le sont moins. Il y a toute une enquête à faire ici, et elle est d'importance, car toutes les représentations ne sont pas susceptibles du même degré de maîtrise.

Par ailleurs, fonder la liberté sur la seule maîtrise potentielle que le sujet peut avoir sur ses représentations est une solution partielle, car elle néglige un élément fondamental de toute prise de décision, à savoir l'intervention de l'humeur, des émotions et des affects. Nous ne rentrerons pas ici dans le débat autour de la question de savoir s'il faut attribuer à chaque émotion un contenu conceptuel-axiologique qui lui servirait en quelque sorte de référent, ou si nous devons envisager qu'il y a dans chaque émotion un noyau qui n'est pas conceptuel². Ce que nous voudrions souligner ici c'est qu'il y a une irréductibilité de la sphère affective et émotive par rapport aux autres formes de contenus représentationnels inclus dans le SRG – notamment les croyances et les valeurs –, et que cette irréductibilité ne permet pas au sujet de

¹ Auquel cas une représentation deviendrait « compréhensive » à partir du moment où elle serait utilisable dans le cadre d'une situation décisionnelle donnée. Nous pensons que cette visée pragmatique n'était pas absente du stoïcisme, et que dans ce contexte aussi la compréhension était là pour soutenir l'action.

² Voir Tappolet (1995) et (2000). Nous nous situons plutôt dans la deuxième optique et nous pensons qu'il y a des éléments qui génèrent dans les émotions et les affects, notamment la *perception* d'un état interne.

s'investir ou de se désinvestir d'un affect ou d'une émotion de la même façon qu'il peut agir sur une de ses valeurs ou de ses croyances. Ce qui ne veut en aucun cas dire qu'il serait plus facile pour le sujet d'arrêter de croire en Dieu que d'arrêter d'avoir peur par exemple. En fait ici ce sont les moyens d'action sur la représentation qui vont différer, et qui vont par là même nous montrer que le concept de liberté n'est peut-être pas univoque.

Si nous nous replaçons dans le cadre de nos hypothèses, cela signifie qu'il y a une forme de liberté attachée à l'évaluation affective, et fondée sur ce que nous pourrions appeler l'*intelligence affective*, ou le bon usage des affects, pour reprendre une image spinoziste. D'autre part, il y a aussi une liberté qui se situe dans le cadre de l'édition doxique. Les deux types de libertés ne sont pas fondés sur les mêmes mécanismes. En effet parvenir à utiliser ou à canaliser ce que le sujet ressent face à une situation donnée, n'est pas du même ressort que d'actualiser telle croyance, telle valeur, ou telle connaissance en vue de la décision dans la même situation. Pourtant les deux aspects vont avoir un effet sur l'ensemble de la prise de décision. Dernier point, si nous avons bien ici deux types de libertés différentes, elles peuvent toutes deux être éduquées dans le sens où le sujet peut apprendre à rectifier ses représentations et à réguler la façon dont il s'investit en elles, parallèlement il doit aussi pouvoir apprendre à mettre à profit le jeu de sa dynamique affective et émotionnelle, au moins partiellement, et apprendre ainsi à intégrer son intelligence affective dans les processus décisionnels.

Le type de liberté qui tourne autour de l'édition axiologique est fondamental, car c'est aussi elle qui va permettre au sujet de se situer face au flux informatif continu qu'il reçoit du monde. *Une certaine finesse dans le choix des représentations extérieures auxquelles le sujet donne son assentiment est nécessaire, notamment dans une société, où du point de vue aussi bien économique que politique, le but est l'orientation du SRD par tout un ensemble de représentations notamment colportées sur le mode de l'image.* Il est clair que si le sujet n'opère pas un tri dans cet ensemble, ses décisions en seront nécessairement affectées¹. Mais il ne faut pas oublier que dans cette dynamique-là, ce sont aussi les affects qui sont visés, et notamment ceux reliés au domaine du désirable au sens large, et du plaisir pris par exemple à la consommation ou à la surconsommation. L'intelligence affective aura donc elle aussi un rôle de premier plan à jouer, par exemple, dans la discrimination qu'elle pourra opérer entre

¹ À cet égard l'exemple de la publicité et de son impact sur le choix des consommateurs pourrait servir à une enquête plus large sur la façon dont notre assentiment est manipulé au quotidien.

les affects qui maximisent réellement la puissance d'agir du SRD et ceux d'une texture plus consumériste.

Il est à noter que ces deux types de libertés peuvent être ramenés à la problématique que nous allons réexaminer maintenant et questionner : le problème de la pondération du présent en contexte décisionnel. Comme nous l'avons vu il existe une forme d'ambivalence dans le rapport que le SRD entretient avec le présent, nous allons nous questionner sur la portée de cette ambivalence et nous interroger sur sa capacité explicative dans des contextes plus étendus.

Une leçon importante à tirer de la Théorie de la Perspective, notamment grâce à la mise au jour des phénomènes de « statu quo bias » et de l'effet de dotation (« endowment effect »)¹ c'est que les choix sont opérés en fonction de l'avoir présent, mais aussi que cet avoir présent peut être surévalué. Les effets liés à la représentation d'une perte possible : aversion au risque dans le domaine des gains et recherche du risque dans le domaine des pertes sont aussi à verser au registre de la façon dont le sujet perçoit son avoir présent. Dans notre travail nous avons montré comment passer du concept d'avoir présent à celui d'état présent afin de pouvoir intégrer dans le fonctionnement du SRD un nombre plus important d'éléments représentationnels. Ce qui ressort de nos analyses, c'est que de la même façon qu'en TP l'avoir présent peut être surévalué, dans le cadre de nos hypothèses, l'état présent du SRD ou du SRG peuvent eux aussi subir une surévaluation, et elle peut avoir diverses conséquences assez significatives.

La tendance conservatrice assez forte du SRD que nous avons mis au jour, à partir des acquis de la TP, tendance à stabiliser la prise de décision habituelle autour de certains axes, est due au rapport qu'entretiennent le SRG et le SRD avec le présent dans les situations non-critiques. C'est justement ce qui fait la texture générale du présent en contexte décisionnel : la plupart des situations que le SRD rencontre quotidiennement sont dans le registre de l'habituel non-critique. Ces situations, de même que certaines situations inhabituelles, permettent au SRD un usage qui est le plus économique et écologique possible. Nous pourrions dire que le décider quotidien se déploie dans une forme de prévisibilité qui permet au SRD de se stabiliser, puisqu'il contrôle très bien les heuristiques habituelles qu'il utilise.

¹ Voir Kahneman/Knetsch/Thaler (1991), in Kahneman/Tversky (Eds.) (2000), p 159-179.

Le décider quotidien s'inscrit par là même dans une certaine forme de *facilité*, que l'on pourrait exprimer ainsi : à partir du moment où le SRD a opté pour une heuristique qui lui semble satisfaisante ou fonctionnelle pour la situation il la transforme en habitude, habitude dont la révision n'arrivera qu'à la condition que sur le court ou le long terme elle manifeste certains défauts évidents, ou bien qu'elle puisse être remplacée par une autre heuristique plus écologique et économique nécessitant encore moins de moyens pour le même résultat ou un résultat supérieur. Comme nous l'avons vu cette tendance conservatrice peut produire une certaine forme d'inertie qui peut à son tour avoir certaines conséquences plus ou moins graves. D'autre part il faut remarquer que plus le sujet évolue dans le temps et plus ses habitudes décisionnelles vont se stabiliser, se sédimenter, et parfois même se scléroser¹.

La procrastination et l'akrasie sont toutes deux à verser au registre d'un certain type de surévaluation pathologique du présent par le SRD. Comme nous l'avons vu, le SRD doit être capable d'évaluer ses propres stratégies, et par là même de les réajuster. Seulement, ce travail de réajustement, surtout dans un contexte habituel non-critique peut ne pas arriver du fait que des pratiques décisionnelles non optimales sont tout à fait compatibles avec le cours normal de l'existence : une stratégie décisionnelle, même si elle ne fonctionne que partiellement dans une certaine situation, du fait de l'habitude, pourra se voir pratiquée indéfiniment, c'est le sens même de l'heuristique que d'être une méthode approximative. Et, hormis dans des cas où l'habitude, qui sera alors qualifiée de mauvaise, peut avoir des conséquences dramatiques pour le sujet, cette sous-optimalité de la stratégie par rapport à la visée n'a qu'une incidence très réduite. Rappelons ici que nous nous situons dans le cadre de situations habituelles non critiques.

Nous rejoignons ici, mais par une autre voie, les hypothèses de Kahneman et Tversky sur les heuristiques et les biais ainsi que celles de Gigerenzer sur la rationalité écologique. Le point étant que dans un contexte tout à fait normal, même si il s'agit d'une expérience, et que le sujet ne sente pas le besoin d'adapter ses stratégies à la situation – si le décideur sait qu'il doit seulement répondre – , il va se comporter de la façon décrite par K/T, c'est-à-dire en négligeant toute une strate de l'information, notamment statistique. Pour la simple raison que

¹ Il y a ici aussi toute une enquête à mener sur le fonctionnement du SRD au travers des différents âges de la vie. Il serait particulièrement intéressant de se pencher sur la question des habitudes décisionnelles, dont on sait par exemple qu'elles sont plus difficiles – voire impossibles – à modifier pendant la vieillesse qu'à toute autre période.

cette information n'a aucune importance pour lui *dans cette situation*. Mais plaçons le même sujet face au même problème dans une situation où l'enjeu est élevé, réellement élevé, par exemple si sa vie ou celle de ses proches dépend directement de la façon dont il va traiter et intégrer l'information statistique dans sa décision, alors nous faisons l'hypothèse que quelle que soit la capacité réelle du sujet à traiter ce type d'information, il lui portera toute son attention. Comme nous le remarquons plus haut, la prise en compte de l'information probabiliste et statistique dépend entièrement du type de situation à laquelle le sujet est confronté. L'important à retenir ici, ce qui nous situe dans une continuité stricte avec Gigerenzer, c'est que dans des situations courantes les sujets n'ont besoin que d'un minimum de rationalité, c'est l'essence même de l'heuristique de reconnaissance par exemple. Dans leur décisions usuelles, les sujets n'utilisent que peu de ressources, car c'est suffisant, et de là à entrer dans un régime d'habitude il n'y a qu'un pas. Il faudrait d'ailleurs analyser de façon plus approfondie et philosophique les rapports entre l'heuristique de reconnaissance et le concept d'habitude, nous pensons que ce dernier pourrait en partie expliquer le premier, car à notre sens la reconnaissance repose sur une forme d'habitude cognitive : les sujets reconnaissent parce qu'ils ont l'habitude de voir ou d'entendre un certain type d'information.

C'est parce que le cours de l'existence décisionnelle – et par là même son inscription dans le présent – permet aux sujets un certain confort habituel que peuvent émerger des comportements tels que la procrastination ou l'akrasie. Ces deux notions sont liées à l'inertie possible que produit le SRD en se stabilisant autour de certaines heuristiques habituelles. Ces comportements possèdent bien entendu divers degrés de gravité selon le domaine dans lequel il se déploie. Ce qui est certain c'est que la plupart des sujets procrastinent à un moment de leur existence ou se trouvent dans une situation akratique¹. Encore une fois tout est une question de degré. Mais la question qu'il convient de se poser pour terminer est la suivante : si dans le cas d'un comportement individuel, les habitudes, une relative inertie, voire même une akrasie ou une procrastination ponctuelles ne prêtent pas à conséquences, qu'en est-il lorsque nous passons de la sphère individuelle aux sociétés ou aux États ? Qu'elles sont les conséquences de ce type de pratiques si nous devons les transposer dans des structures plus vastes impliquant des enjeux plus élevés ?

¹ C'est ce que soulignent entre autres les articles d'Akerlof (1991), Steel (2007), Ainslie (2005) et (2010).

Comme nous l'avons vu plus haut, le rôle de la phase d'évaluation dans le SRD est aussi d'anticiper les conséquences possibles des alternatives qui peuvent être choisies, même si cette anticipation peut être parfois vague ou nébuleuse du fait de l'opacité de l'avenir en tant que tel, hormis pour des situations encore une fois très claires, très déterminées ou très maîtrisées. Anticiper les conséquences potentiellement néfastes de ses choix est une des fonctions du SRD. Les questions que nous aimerions soulever ici sans pouvoir les traiter dans toute leurs extensions sont les suivantes : peut-on attribuer à des structures plus grandes et plus complexes que de simples individus un SRD ? Cette attribution permettrait-elle par exemple d'expliquer le comportement des États face à des enjeux majeurs par exemple la crise climatique en terme de procrastination ? Une procrastination dont les conséquences seraient pour le coup beaucoup plus graves et générales que dans le décider quotidien d'un sujet isolé.

En ce qui concerne la première question, nous donnerons les éléments de réponse suivants : à partir du moment où une structure a pour vocation de prendre des décisions et que ces décisions se situent dans le registre de la représentation, il y a de fortes chances, que les processus décisionnels aient la même forme que dans le SRD singulier à savoir : visée, édition et évaluation. La différence majeure va se jouer dans le *régime représentationnel* dans lequel ces processus prennent place. Si la forme est la même, et si nous nous situons par exemple dans le cadre d'un État, c'est au niveau du contenu des représentations qui vont entrer en compte dans les décisions que les choses seront différentes. Dans le domaine politique, l'édition des alternatives se fait *normalement* à partir d'un traitement exhaustif de l'information disponible, traitement qui doit aussi faire autant que possible l'économie de l'édition affective des alternatives. En effet, nous pouvons penser que dans le cadre des décisions étatiques, ce sont normalement des experts qui construisent les alternatives et les évaluent, édition et évaluation qui doivent se faire en continuité avec les préférences générales exprimées par les citoyens. Si l'édition affective des alternatives est *normalement* évitée, car elle rendrait l'opération partielle et partielle, l'édition doxique est opérée de diverses façons puisque le pouvoir qui assume la prise de décision n'est jamais neutre politiquement. De façon assez schématique nous pouvons dire que l'édition doxique des alternatives correspondra en gros à la mouvance politique dans laquelle s'inscrit le pouvoir en place, mais aussi, et c'est très important, aux influences diverses et inévitables que le pouvoir en place subit de façon constante.

Comme nous venons de le souligner une des différences majeures entre décision individuelle et décision étatique se situe dans l'édiction affective qui est *normalement* remplacée par l'expertise. Cette obligation d'expertise entraîne une autre différence d'avec la décision subjective : le processus décisionnel n'est plus alors produit par un seul sujet, mais par un ensemble de sujets en situation de coopération décisionnelle, sujets qui sont autant d'experts. Cette coopération, qui est très loin d'aller de soi, puisqu'elle implique une pluralité de sujets, donne à la décision politique une texture particulière. Une texture dont on pourrait dire qu'elle est presque nécessairement « impersonnelle », même si elle peut s'inscrire dans une ligne politique très claire et très précise, ce qui la rend par contre alors identifiable.

Pour les décisions étatiques, qui peuvent elles aussi être soumises aux différents types de situations analysées dans le chapitre trois, comme l'expertise est censée être maximale, l'évaluation des conséquences potentielles des alternatives éditées doit être, ou *devrait être*, elle aussi maximale. *Idéalement* les États ne devraient pas développer de comportements akratiques ou procrastinateurs, à moins que ce ne soit pour le bien du plus grand nombre. La question que nous devons nous poser face à certaines situations globales contemporaines est la suivante : peut-on attribuer aux États des comportements de type akratiques ou procrastinateurs ? Et, si la réponse est positive, à quoi ces comportements doivent-ils être attribués ?

La situation écologique mondiale¹, dont la crise climatique est un épiphénomène, de par la forme qu'elle prend, et parce qu'elle concerne un ensemble d'États dont la majeure partie – par exemple les États membres du G 20 – sont surinformés, nous offre un exemple problématique de ce qui semble être un comportement akratique *et* procrastinateur à très grande échelle. Akratique, car le « meilleur » est parfaitement vu (« video meliora »), parfaitement connu, compris, mais il n'est pas choisi ; procrastinateur, car la mise en place de solutions étatiques fiables semble toujours reportée dans un futur problématique, un futur dont le postulat implicite – tel que le souligne notamment Hans Jonas² – est qu'il y aura toujours

¹ Nous subsumons sous ce concept certes un peu vague la totalité des pratiques humaines qui ont un impact destructeur sur les conditions de possibilité et de continuation de la vie elle-même. Ces pratiques peuvent être volontaires (déforestation, pêche massive, etc.) ou pas (accidents nucléaires), et avoir des effets directs ou indirects, par exemple sur des échelles temporelles variées allant de l'instantanéité à la demi-vie atomique...

² Voir Hans Jonas, *Le principe responsabilité*. Jonas associe avec raison la situation présente avec la transformation de l'agir humain, transformation qui se manifeste par la possibilité pour l'homme, inédite, de réellement modifier le cours des cycles naturels, voir pages 22-62. Pour le concept d'humanité future, voir pages 87-99.

une humanité demain ou une autre génération pour résoudre les problèmes. Mais l'ampleur de la crise est telle que depuis quelques décennies elle semble remettre en question la validité de ce postulat, qui passe pour le coup au rang d'hypothèse... Nous n'irons pas plus loin dans la description factuelle, elle est connue. Nous voulons ici confronter ces phénomènes à notre théorie, en nous demandant si elle permet de les comprendre ou de les expliquer.

Comme nous l'avons vu, au niveau individuel, la procrastination est due à une surévaluation de la valeur ou de l'utilité de la situation présente, au détriment de la réalisation d'une action qui *doit* être faite, mais est indéfiniment repoussée. L'akrasie, bien que participant aussi de cette surévaluation de l'état présent, est à relier au phénomène d'aversion au risque et surtout au fait que les pertes pèsent plus lourd que les gains. La question est : que se passe-t-il lorsque nous passons à des niveaux plus généraux ? Au niveau étatique par exemple.

Ici nous avons deux options, ou nous choisissons de penser que l'akrasie et la procrastination étatique sont seulement le reflet et l'agrégation des comportements individuels des citoyens qui composent les États, un peu à la façon dont une préférence collective ou une tendance peut émerger d'un ensemble de sujets¹. C'est une possibilité mais elle est partielle parce qu'elle postule que la structure étatique est totalement perméable, transparente, aux comportements individuels, ce dont nous doutons. Elle a aussi le tort de renvoyer la responsabilité de la situation actuelle *seulement* sur les citoyens. La deuxième solution est plus compliquée à penser, plus nuancée, elle consiste à dire qu'il y a certes une influence probable du collectif sur l'État, mais que comme ce dernier est soumis à un *régime représentationnel* spécifique qui recoupe à certains niveaux – notamment au niveau des valeurs communes, des croyances, etc. – celui des individus, il doit être doté d'un SRD spécifique, et que ses comportements akratiques et procrastinateurs lui sont alors propres. C'est la voie qui nous semble la plus pertinente, car elle permet de composer l'influence des comportements individuels avec la spécificité du comportement étatique, elle pose toutefois un problème majeur : dans le cadre de nos hypothèses comment comprendre l'akrasie et la procrastination étatiques ? Car le passage du particulier vers l'État ne va pas de soi, et nous ne pourrions qu'aborder le problème.

¹ Voir à ce sujet, Arrow (1963) et Picavet (1996), p 269-380, qui commente d'ailleurs magistralement le texte d'Arrow. Nous ne rentrerons pas ici dans la problématique capitale des préférences générales et de la « volonté » générale.

Pour comprendre comment ces phénomènes sont possibles dans la sphère étatique il faudrait analyser de façon très fine le *régime représentationnel* à partir duquel les décisions politiques sont prises. Ces décisions, comme les décisions individuelles, ne peuvent être désolidarisées d'un ensemble de référence qui est un ensemble de représentations. Même si il est bien clair que les représentations au niveau étatique ne se donnent pas sur le même mode et la même forme qu'au niveau individuel, elles ont quand même un effet déterminant sur la qualité des décisions. Les éléments pour parvenir à comprendre l'akrasie et la procrastination au niveau étatique vont donc eux aussi se situer à *l'articulation* du système de représentation sur lequel le pouvoir se fonde et du SRD. Par exemple l'influence de certains éléments représentationnels d'ordre économique comme la toute-puissance du marché, ou du « Prométhée définitivement déchaîné »¹ dans la sphère de l'hyper technicisation permettent de comprendre comment l'édiction des alternatives dans le domaine politique peut être influencée, voire déterminée.

Si il existe une procrastination et une akrasie de type étatique, ce qui semble être le cas, doivent-elles être elles aussi ramenées à une surévaluation du présent ? Comme dans la sphère individuelle. À notre sens oui, mais il semblerait que les choses soient plus complexes que dans le décider individuel. Car ce qu'il faut arriver à penser dans le cadre de systèmes plus importants c'est le fait qu'ils ne sont pas soumis à *un seul régime représentationnel* comme dans le cas individuel, mais à une *multitude de régimes*, parfois incompatibles les uns avec les autres, et que ces régimes sont autant d'*influences* sur les décisions. Dans le domaine de la décision individuelle le concept d'*influence*, bien qu'ayant une certaine importance, notamment parce que les éléments du SRG influent le travail du SRD, n'a pas été construit de façon systématique. Dans le cadre de décisions politiques c'est un concept clé, car il est de l'essence des États modernes d'être soumis à un régime d'influences qui est aussi un régime de représentations, à notre sens ce régime entre aussi dans l'explication de certains comportements akatriques et procrastinateurs à grande échelle, car les influences ou les pressions que subit la sphère politique, par exemple de la part de l'économie et des lois du marché, permet dans une large mesure d'expliquer des faits aussi graves que les déforestations massives, ainsi que l'usage continué d'énergies polluantes.

¹ *Le principe responsabilité*. Page 1. Voir aussi le très pertinent et très beau texte de Saint-Sernin (1973), p 258.

Dans un domaine où la prévision sur le long terme devrait être la règle d'or pour le bien du plus grand nombre il est pour le moins paradoxal de se retrouver face à une surévaluation du présent qui met en danger la pérennité de l'espèce humaine elle-même. Surévaluation dont une des raisons principales est que dans la construction de leurs décisions les États modernes obéissent à la politique du profit et à la toute-puissance de l'économie. Si nous faisons l'hypothèse qu'un système de référence décisionnel existe dans la sphère politique, nous assistons aujourd'hui à une *défaillance majeure de la phase d'évaluation*. Mais il faut bien prendre garde que cette défaillance ne doit pas être versée au registre d'un manque de connaissance, ou d'une construction biaisée des alternatives, au contraire, tout est parfaitement su et conçu, et cette défaillance est plutôt à relier à deux éléments qui viennent court-circuiter l'ensemble du processus décisionnel : *la surévaluation du présent parce que c'est le temps du profit et des affaires, la faillite des États à proposer des projets de vie et de sociétés*. À partir du moment où le SRD politique oscille de façon habituelle entre ces deux pôles, la plupart du temps de façon tout à fait consciente, et qu'il subit de sucroît tout un ensemble d'influences auquel il ne peut guère échapper, il n'est pas étonnant de le voir développer des comportements dont on pourrait dire alors qu'ils sont pour le coup réellement paradoxaux et irrationnels.

Nous pourrions dire qu'aujourd'hui ce n'est plus dans le sujet singulier qu'il faut chercher les traces ou la marque d'une irrationalité constitutive, presque essentielle, même si ce sujet est en puissance de mal construire et mal composer ses décisions. À cet égard nous devons beaucoup aux travaux de Gigerenzer et de son groupe de recherche. *C'est dans les décisions étatiques que se jouent aujourd'hui la réelle irrationalité, la réelle et menaçante incohérence qui s'inscrit dans la réalité comme une akrasie et une procrastination généralisées à la limite du compréhensible et du pensable.*

Pourquoi les êtres humains décident-ils ? Parce qu'ils ont des projets, sur le court ou le long terme, des projets importants ou modestes. Or nous observons dans la sphère politique une carence béante dans la capacité qu'elle devrait avoir à proposer des projets, c'est-à-dire à proposer des visées, à faire de l'avenir un but commun. La politique a perdu une grande partie de sa pertinence à partir du moment où elle a renoncé à faire de l'homme un projet d'avenir.

L'opération d'établissement de la visée par le SRD, qui peut paraître presque anodine tellement elle est courante, révèle pourtant un élément essentiel de tout être humain qui est de

se construire comme projet, comme à-venir, comme à-faire et cette construction de l'homme par lui-même passe par l'agir, c'est-à-dire par un décider préalable dont la réalisation, l'inscription dans le monde sur le mode du faire, ou du dire, sur n'importe quel mode, est essentielle à la construction de l'humanité de l'humain. Cette « atmosphère d'humanité » dont parlait Merleau-Ponty n'est possible que parce que les hommes laissent leurs traces à la surface du monde. Un monde n'est humain que dans l'espace possible qu'il laisse à l'inscription du décider, et dans l'espace ouvert de la visée comme chose à faire, comme projet à commencer et à recommencer.

Bibliographie

- Abecassis, (1999) : « L'interaction médecin - patient : quelques enseignements de la théorie des jeux », in Audier, F., Gazier, B., Outin, J.L. (Éds.) (1999) : *L'Économie sociale : formes d'organisations et institutions*, Paris, L'Harmattan.
- Ainslie, G. (2005) : « Précis of *breakdown of Will* ». *Behavioral and Brain Sciences*. Vol. 28, p 635-673.
- Ainslie, G. (2010) : « Procrastination. The basic impulse ». In Andreou, C., White, M. (Eds.) (2010), p 11-27.
- Akerlof, G. A. (1991) : « Procrastination and Obedience ». *The American Economic Review*. Vol. 81, Issue 2, p 1-19.
- Allais, M. (1952) : « The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School ». In Allais, M., Hagen, O. (Eds) (1979), p 25-145.
- Allais, M. (1953) : « Le Comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine », *Econometrica*. Vol. 21, p. 503-546
- Allais, M., Hagen O. (1979) : *Expected Utility Hypotheses and the Allais Paradox*. D. Reidel Publishing Company.
- Amihud, Y. (1979) : « A Reply to Allais ». In Allais, M., Hagen, O. (Eds) (1979), p 185-190.
- Amihud, Y. (1979) : « Critical Examination of the New Foundation of Utility ». In Allais, M., Hagen, O. (Eds) (1979), p 149-160.
- Anand, P., Pattanaik, P. K., Puppe, C. (Eds.) (2009) : *The handbook of rational and social choice*. Toronto : Oxford University Press.
- Andler, D. (2004) : *Introduction aux sciences cognitives*. Gallimard, Folio Essais.
- Andreou, C. White, M. (Eds) (2010) : *The Thief of Time. Philosophical Essays on Procrastination*. Oxford University Press.
- Angrilli, A., Cherubini, P., Pavese, A., Manfredini, S. (1997) : « The influence of affective factors on time perception ». *Perception & Psychophysics*. Vol. 59, p 972-982.
- Anscombe, F. J., Aumann, R. J. (1963) « A Definition of Subjective Probability ». *The Annals of Mathematical Statistics*. Vol. 34, No. 1, p 199-205.
- Ariely, D., Zakay, D. (2001) : « A timely account of the role of duration in decision making ». *Acta Psychologica*. Vol. 108, p 187-207.
- Ariely, D., Wertenbroch. K. (2002) : « PROCRASTINATION, DEADLINES, AND PERFORMANCES : Self-Control by Precommitment. *Psychological Science*. Vol. 13, No. 3, p 219-224.
- Aristote, *Éthique à Nicomaque*. Traduction française R. Bodéüs. GF Flammarion, 2004.
- Arnauld et Nicole (1662) : *La logique ou l'art de penser*. Tel, Gallimard, 1992.
- Asgary, A., Levy, J. (2009) : « A Review of the Implications of Prospect Theory for Natural Hazards and Disaster Planning ». *Int. J. Environ. Res*. Vol. 3, No. 3, p 379-393.
- Aumann, R. J., Hart, S., Perry, M. (2005) : « Conditioning and the Sure-Thing Principle ». *The Hebrew University of Jerusalem, Center for the Study of Rationality*. No. 393, p 1-10.
- Avni-Babad, D., Ritov, I. (2003) : « Routine and the perception of time ». *Journal of Experimental Psychology*. Vol. 132, No. 4, p 534-550
- Baird, D. (1992) : *Inductive Logic*. Prentice Hall. New Jersey.
- Bar-Hillel, M. (1973) : « Studies of Representativeness. » In Kahneman, D., Slovic, P., Tversky, A. (eds.) (1982), p 69-83.
- Barbaras, R. (2008) : *Introduction à une phénoménologie de la vie*. Vrin, Problèmes et controverses.
- Barberà, S., Hammond, P. J., Seidl, C. (Eds.) (2004) : *Handbook of Utility Theory*. Volume 2, Extensions. Kluwer Academic Publishers.

- Baron, J. (1988) : *Thinking and Deciding*. Reprinted in 2009. Cambridge University Press.
- Bayes, T. (1763) : « An Essay toward solving a Problem in the Doctrine of Chances ». *Philosophical Transactions of the Royal Society of London*. Vol. 53, p 370-418.
- Benoist, J. (Dir.) (2008) : *La conscience du temps. Autour des Leçons sur le temps de Husserl*, Paris, Vrin, collection « Problèmes et controverses ».
- Bernard, L. C., Mills, M., Swenson, L., Walsh, R. P. (2006) : « An Evolutionary Theory of Human Motivation ». *Genetic, Social, and General Psychology Monographs*. Vol. 131, No. 2, p 129-184
- Bernheim, B. D., Rangel, A. (2009) : « Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics ». *Quarterly Journal of Economics*. Vol. 124, No. 1, p 51-104.
- Bernoulli, D. (1738) : « Exposition of a New Theory on the Measurement of Risk ». English translation 1954. *Econometrica*. Vol. 22, No. 1, p 23-36.
- Bernoulli, J. (1713) : *Ars conjectandi*. Basel.
- Bernstein, P. L. (1998) : *Against the Gods : a Remarkable Story of Risk*. John Wiley & Sons.
- Bettman, J. R., Luce, M. F., Payne, J. W. (1998) : « Constructive Choice Consumer ». *Journal of Consumer Research*. Vol. 25, No. 3, p 187-217.
- Binmore, K. (2009) : *Rational Decisions*. Princeton University Press.
- Binmore, K., Stewart, L., Voorhoeve, A. (2012) : « How much Ambiguity Aversion ? Finding Indifferences between Ellsberg's Risky and Ambiguous Bets ». *Journal of Risk and Uncertainty*. Vol. 45, No. 3, p 215-238.
- Birnbaum, M. H. (2004) : « Causes of Allais common consequence paradoxes: An experimental dissection ». *Journal of Mathematical Psychology*. Vol. 48, Issue 2, p 87-106.
- Birnbaum, M. H. (2006) : « Evidence against prospect theories in gambles with positive, negative, and mixed consequences ». *Journal of Economic Psychology*. Vol. 27, p 737-761.
- Birnbaum, M. H., Schmidt, U. (2010) : « Allais Paradoxes Can be Reversed by Presenting Choices in Canonical Split Form ». *Working Paper*. Kiel Institute for the World Economy. No. 1615.
- Blaszczynski, A., Nower, L. (2001) : « A pathways model of problem and pathological gambling ». *Addiction*. Vol. 97, p 487-499.
- Bovens, L., Hartmann, S. (2003) : *Bayesian Epistemology*. Oxford University Press.
- Bradley, R. (2004) : « Ramsey's Representation Theorem ». *Dialectica*, Vol. 58, No. 4, p 483-498.
- Bradley, R. (2006) : « A Unified Bayesian Decision Theory ». *CPNSS Working Paper*. Vol. 1, No. 1, p 1-18.
- Bradley, R. (2007) : « Ramsey and the measurement of belief. In Corfield, D., Williamson, J. (eds.) : *Foundations of Bayesianism*. Applied logic series. Kluwer Academic Publishers, London, UK, p. 263-290
- Branas-Garza, P., Georgantzis, N., Guillen, P. (2007) : « Direct and indirect effects of pathological gambling on risk attitudes ». *Judgment and Decision Making*. Vol. 2, No. 2, p 126-136.
- Buehler, R., Griffin, D., Ross, M. (2002) : « Inside the Planning Fallacy : The Causes and Consequences of Optimistic Time Predictions ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 250-251.
- Cadrin, G. R. (2011) : *L'être humain face à lui-même. L'animal autoréférent*. L'Harmattan, Ouverture philosophique.

- Camerer, C. F., Lovo, D. (1999) : « Overconfidence and Excess Entry ». *American Economic Review*. Reprinted in Kahneman, D., Tversky, A. (Eds.) (2000), p 414-423.
- Carnap, R. (1966) : *Philosophical foundations of physics*. New York, Basic Books.
- Chase, V. M., Hertwig, R., Gigerenzer, G. (1998) : « Visions of Rationality ». *TRENDS in Cognitive Science*. Vol. 2, No. 6, p 206-214.
- Chater, N., Oaksford, M. (Eds.) (2008) : *The Probabilistic Mind*. Oxford University Press.
- Chernoff, L. E. M. (1959) : *Elementary Decision Theory*. Dover.
- Choquet, G. (1955) : « Theory of capacities ». *Annales de l'Institut Fourier*, 5, 131-295.
- Colman, T. S. (2011) : « Probability, Expected Utility, and the Ellsberg Paradox ». *Working Paper*. D8, D81. University of Chicago - Becker Friedman Institute for Research in Economics; Close Mountain Advisors LLC.
- Coombs, C. H., Milholland, J. E. (1954) : « Testing the 'rationality' of an individual's decision making under uncertainty ». *Psychometrika*.
- Coombs, C. H., Dawes, R. M., Tversky, A. (1970) : *Mathematical Psychology*. Englewood Cliffs, N.Y., Prentice-Hall
- Cosmides, L., Tooby, J. (1996) : « Are humans good intuitive statisticians after all ? Rethinking some conclusions from the literature on judgment under uncertainty ». *Cognition*. Vol. 53, t. 73.
- Cosmides, L., Tooby, J. (2000) : « Evolutionary Psychology and Emotions ». In Lewis, M., Haviland-Jones, J. M., (Eds.) (2000) : *Handbook of Emotions*. 2nd Edition, NY : Guilford.
- Coumet, E. (1970) : « La théorie du hasard est-elle née par hasard ? ». *Annales. Économies, Société, Civilisations*. 25e année, No. 3, p 574-598.
- Damasio, A. (1994) : *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam Publishing.
- Daston, L. (1988) : *Classical Probability in the Enlightenment*. Princeton University Press.
- Daston, L. (1989) : « L'interprétation classique du calcul des probabilités ». *Annales. Économies, Sociétés, Civilisations*. Vol. 44, No. 3, p 715-731.
- David, F. N. (1955) : « Dicing and gambling ». In Pearson, E. S., and M G Kendall, M. G. (Eds.) (1955).
- Davidson, D., Suppes, P. (1956) : « A Finistic Axiomatization of Subjective Probability and Utility ». *Econometrica*. Vol. 24, No. 3. p 264-275.
- Davidson, D. (1982) : « Paradoxes of irrationality ». In Wollheim, R., Hopkins, J. (Eds.) (1982) : *Philosophical Essays on Freud*. Cambridge University Press.
- Davidson, D. (2001) : *Essays on Actions and Events*. Second Edition. Clarendon Press. Oxford.
- de Finetti, B. (1937) : « La Prévision : ses lois logiques, ses sources subjectives ». *Annales de l'I.H.P.* Tome 7, No. 1, p1-68.
- de Finetti, B. (1989) : « Probabilism. A Critical Essay on the Theory of Probability and on the Value of Science ». *Erkenntnis*. Vol. 31, p 169-223.
- De Sousa, R. (1987) : *The rationality of emotion*. Cambridge, Mass. : MIT Press
- Dennett, D. (1978) : *Brainstorms : Philosophical Essays on Mind and Psychology*. Montgometry, Vt.: Bradford Books.
- Dickes, P., Tournois, J., Flieller, A., Kop, J-L. (1994) : *La psychométrie : théories et méthodes de la mesure en psychologie*. Paris: PUF.
- Dokic, J., Engel, P. (2002) : *Frank Ramsey truth and success*. London : Routledge.
- Droit-Volet, S., Meck, W. (2007) : « How emotions colour our perception of time ». *TRENDS in Cognitive Sciences*. Vol. 11, No. 12, p 504-513.

- Dumezadier, J. (1973) : *Sociologie empirique du temps libre*. Le Seuil.
- Eagleman, D. M. (2008) : « Human time perception and its illusions » *Current Opinion in Neurobiology*. Vol. 18, p 131-136.
- Edwards, W. (1954) : « The Theory of Decision making ». *Psychological Bulletin*. Vol. 51, No. 4, p 380-417.
- Edwards, W. (1961) : « Behavioral Decision Theory ». *Annual Review of Psychology*. Vol. 12, p 473-498.
- Edwards, W. (1962) : « Dynamic decision theory and probabilistic information processing ». *Human Factors*. Vol. 4, p 59-73.
- Edwards, W., Lindman, H., Savage, L. J. (1963) : « Bayesian Statistical Inference for Psychological Research ». *Psychological Review*. Vol. 70, p 193-242.
- Elliot, A. J., Dweck, C. S. (Eds.) (2005) : *Handbook of competence and motivation*. Guilford Press.
- Ellsberg, D. (1961) : « Risk, ambiguity, and the Savage axioms ». *Quarterly Journal of Economics*. Vol. 75, p 643-669. Reprinted in Gärdenfors, P., Sahlin, N-E. (Eds.) (1988), p 245-269.
- Engel, P. (1981) : « Davidson en perspective ». *Critique*, 409-410, p 578-594.
- Engel, P. (1983) : « Comment fixer la croyance ». *Critique*. 432, p 384-408.
- Engel, P. (1984) : « Aristote, Davidson et l'akrasia », *Philosophie*. Vol. 3.
- Engel, P. (1995) : « Les croyances », in D. Kambouchner, (Ed.) (1995). *Les notions philosophiques*, Paris, Gallimard, 1995, tome II, 1-101
- Engel, P. (1997) : « Le rôle de la croyance dans l'explication de l'action », in Petit, J-L. (Ed.) : *Les neurosciences et la philosophie de l'action*, Paris, Vrin 1998, p 327-339.
- Engel, P. (Dir.) (2000) : *Précis de philosophie analytique*. Thémis, Philosophie, P.U.F
- Engel, P. (2001) : « Sommes-nous responsables de nos croyances ? ». In Michaud, Y. (Ed.) : *Université de tous les savoirs, vol. VI. Qu'est-ce que la culture ?* O. Jacob, p 429-439.
- Engel, P. (2002) : « La rationalité est-elle incodifiable ? ». *Philosophia Scientiae*. Vol. 6, No. 2, p 91-107.
- Engel, P. (2005) : « Belief as a disposition to act. Variation on a pragmatist theme ». *Cognitio, Centro de Estudios pragmaticos, Universidade catolica de San Paulo*.
- Engel, P. (2005) : « Tacit Belief ». *Synergies*. Center of Advanced Studies, Oslo.
- Engel, P. (2007) : « Akrasia Pratique et akrasia épistémique », *Le philosophoire*, n°19, "L'Action".
- Engel, P. (2007) : « Belief and Normativity ». *Disputatio*, Lisboa, 2008, p 153-177.
- Epstein, L.G., Le Breton, M. (1993) : « Dynamically Consistent Beliefs must be Bayesian ». *Journal of Economic Theory*. Vol. 61, p 1-22.
- Etner, J., Jeleva, M., Tallon, J-M. (2010) : « Decision under Ambiguity ». *Journal of Economics Survey*. Vol. 26, No. 2, p 234-270.
- Fenouillet, F. (2009) : La « cagnotte » et les théories de la motivation. *Le Cercle de Recherche et d'Action Pédagogiques et les Cahiers pédagogiques*. Revue en ligne.
- Ferreri, M., Godefroi, M., Nuss, P., Ferreri, F. (2002) : « Dysmorphophobie, Peur de dysmorphie, dysmorphesthésie ». *Ecyelopédie Médico-chirurgicale*. 37-146-A-10.
- Fienberg, S. E., Zellner, A. (Eds.) (1975) : *Studies in Bayesian Econometrics and Statistics*. North-Holland / American Elsevier.
- Fink, E. (1960) : *Le jeu comme symbole du monde*. Traduction française Hildenbrand et Al. (1966). Les Éditions de Minuit.
- Fishburn, P. C. (1970) : *Utility Theory for decision making*. New York ; Wiley, Toronto.

- Fishburn, P. C. (1973) : *Les mathématiques de la décision*. Paris, Mouton.
- Fishburn, P. C. (1986) : « The Axioms of Subjective Probability ». *Statistical Review*. Vol. 1, No. 3, p 335-358.
- Fishburn, P., Wakker, P. (1995) : « The Invention of the Independence Condition for Preferences ». *Management Science*. Vol. 41, No. 7, p 1130-1143.
- Fischhoff, B. (2002) : « Heuristics and Biases in Application ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 730-748.
- Fox, C. R., Tversky, A. (1995) : « Ambiguity Aversion and Comparative Ignorance ». *Quarterly Journal of Economics*. Vol. 110, No. 3, p 585-603.
- Fox, C. R., Tversky, A. (1998) : « A Belief-Based Account of Decision under Uncertainty ». *Management Science*. Vol. 44, No. 7. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 118-142.
- Frankfurt, H. G. (1978) : « The Problem of Action ». *American Philosophical Quarterly*. Vol. 5, No. 2, p 157-162.
- Frederick, S. (2002) : « Automated Choice Heuristic ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 548-558.
- Friedman, M., Savage, L. J. (1948) : « The Utility Analysis of Choice Involving Risk ». *The Journal of Political Economy*. Vol. LVI, No. 4, p 279-303.
- Friedman, M., Savage, L. J. (1952) : « The expected-utility hypothesis and the measurability of utility ». *The Journal of Political Economy*. Vol. LX, No. 6, p 463-473.
- Galavotti, M. C. (1989) : « Anti-Realism in the Philosophy of Probability : Bruno de Finetti's Subjectivism ». *Erkenntnis*. Vol. 31, p 239-261.
- Galavotti, M.C. (1991) : « The notion of subjective probability in the work of Ramsey and de Finetti ». *Theoria*. Vol. LXII, No. 3, p 239-259.
- Gärdenfors, P., Sahlin, N-E. (Eds.) (1988) : *Decision, Probability, and Utility*. Cambridge University Press.
- Gigerenzer et Al. (1989) : *The Empire of Chance*. Cambridge University Press.
- Gigerenzer, G. (1991) : « How to Make Cognitive Illusions Disappear : Beyond 'Heuristics and Biases' ». in Stroebe, W., Hewstone, M. (Eds.). (1991). *European Review of Social Psychology* (Vol. 2, pp. 83-115). Chichester: Wiley.
- Gigerenzer, G. (1993) : « The bounded rationality of probabilistic mental models ». In Manktelow, K. I., Over, D. E. (Eds.) (1993) : *rationality : Psychological and philosophical perspective*. London : Routledge.
- Gigerenzer, G. (1996) : « On Narrow Norms and Vague Heuristics : A Reply To Kahneman and Tversky ». *Psychological Review*. Vol. 103, No. 3, p 592-596.
- Gigerenzer, G., Goldstein, D. G. (1996) : « Reasoning the Fast and Frugal Way : Models of Bounded Rationality ». *Psychological Review*. Vol. 103, p 650-669. Reprinted in Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011), p 33-55.
- Gigerenzer, G., Todd, P. M., and the ABC Research Group (Eds.) (1999) : *Simple Heuristics that Make Us Smart*. Oxford University Press.
- Gigerenzer, G., Selten, R. (Eds) (2001) : *Bounded Rationality. The Adaptive Toolbox*. The M.I.T Press.
- Gigerenzer, G. (2004) : « Fast and Frugals Heuristics : The Tools of Bounded Rationality ». In Koehler, D. J, Harvey N. (Eds.) (2004), p 62-89.
- Gigerenzer, G. (2008) : *Rationality for Mortals*. Oxford University Press.
- Gigerenzer, G., Brighton, H (2009) : « *Homo heuristicus* : why Biased Minds Make Better Inferences ». *Topics in cognitive Science*. Vol. 1, p 107-143. Reprinted in Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011), p 2-27.

- Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011) : *Heuristics. The Foundations of Adaptive Behavior*. Oxford University Press.
- Gilbert, D. T., Wilson, T. D. (2000) : « Miswanting : Some Problems in the Forecasting of Future Affective States. » In Forgas, J. (Ed.) (2000) : *Feeling and thinking : the role of affect in social cognition*. Cambridge University Press. Reprinted in Lichtenstein, S., Slovic, P. (Eds.) (2006), p 550-563.
- Gilboa, I., Postlewaite, A., Schmeidler, D. (2004) : « Rationality of Belief or : Why Bayesianism is neither Necessary nor Sufficient for Rationality ». *Working Paper No. 14*, The Foerder Institute for Economic Research
- Gilboa, I. (2009) : *Theory of Decision under Uncertainty*. Cambridge University Press.
- Gilboa, I., Postlewaite, A., Schmeidler, D. (2009) : « Is It always Rational to Satisfy Savage's Axioms ? ». *Economics and Philosophy*. Vol. 25, p 285-296.
- Gilboa, I. (2011) : *Making Better Decisions*. Wiley-Blackwell.
- Gillies, D. (1998) : « The Duhem Thesis and the Quine Thesis », in Curd, M. and Cover, J. A. (Eds.) : *Philosophy of Science: The Central Issues*. New York: Norton (1998), p 302-319.
- Gillies, D. (2000) : *Philosophical theories of probability*. Routledge, London and New York.
- Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002) : *Heuristics and Biases*. Cambridge University Press.
- Gilovich, T., Griffin, D. (2002) : « Heuristics and Biases : Then and Now ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 1-18.
- Gintis, H. (2009) : *The bounds of reason : game theory and the unification of behavioral sciences*. Princeton University Press.
- Goddard, J. C., Labrune, M., Barbaras, R. (1992) : *Le corps*. Vrin, Intégrale.
- Goldstein, D. G., Gigerenzer, G (2002) : « Models of Ecological Rationality: The Recognition Heuristic». *American Psychological Review*. Vol. 109, p 75-90. Reprinted in Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011), p 60-81.
- Goldstein, W. M., Einhorn, H. J. (1987) : « Expression Theory and the Preference Reversal Phenomena ». *Psychological Review*. Vol. 94, No. 2, p 236-254.
- Gonzalez, C. (2004) : « Decision support for real-time decision-making tasks ». *Organizational Behavior and Human Decision Processes*. Vol. 96, p 142-154.
- Granger, G. G. (1967) : *Pensée formelle et sciences de l'homme*. Archives Karéline, Bookollegium. Fac-similé (2010).
- Granger, G. G. (1988) : *Essai d'une philosophie du style*. Odile Jacob.
- Granger, G. G. (1995) : *Le probable, le possible et le virtuel*. Odile Jacob.
- Grether, D. M., Plott, C. R. (1979) : « Economic Theory of Choice and the Preference Reversal Phenomena ». *American Economic Review*. Vol. 69, p 623-628. Reprinted in Lichtenstein, S., Slovic, P. (Eds.) (2006), p 77-93.
- Griffiths, T. L., Yuille, A. (2008) : « A primer on probabilistic inference ». In Chater, N., Oaksford, M. (eds.) (2008), p 33-57.
- Hacking, I (1967) : « Slightly more realistic personal probability ». *Philosophy of Science*. Vol. 34, p 311-325. Reprinted in Gärdenfors, P., Sahlin, N-E. (Eds.) (1988), p 119-135.
- Hacking, I. (1971) : « Equipossibility Theories of Probability ». *The British Journal for the Philosophy of Science*. Vol. 22, No. 4, p 339 -355. Oxford University Press.
- Hacking, I. (1975) : *The Emergence of Probability*. Second Edition 2006. Cambridge University Press.
- Hacking, I. (1990) : *The Taming of Chance*. Twelfth Printing (2009). Cambridge University Press.

- Hacking, I. (2001) : *L'ouverture au probable. Éléments de logique inductive*. Traduction française Dufour, M. (2004). Armand Colin.
- Halevy, Y. (2007) : « Ellsberg Revisited : An Experimental Study » *Econometrica*. Vol. 75. No. 2, p 503-536.
- Hammond, J. S., Keeney, R. L., Raiffa, H. (1999) : *Smart Choices*. Broadway Books, New York.
- Hancock, P. A., Block, R. A. (2012) : « The Psychology of Time : A View Backward and Forward ». *American Journal of Psychology*. Vol. 125, No. 3, p 267-274.
- Hansen, J., Trope, Y. (2012) : « When Time Flies : How Abstract and Concrete Mental Construal Affect the Perception of Time ». *Journal of Experimental Psychology*. Advance Online Publication. Doi : 10.1037/a0029283.
- Hansson, B. (1988) : « Risk aversion as a problem of conjoint measurement ». In Gärdenfors, P., Sahlin, N-E. (Eds.) (1988), p 136-161.
- Hare, R. M. (1952) : *The Language of Morals*, Oxford: Clarendon Press.
- Hare, R. M. (1963) : *Freedom and Reason*, Oxford: Clarendon Press.
- Hausner, M. (1954): « Multidimensional utilities ». In Thrall, R.M. et Al. (Eds.) : *Decision Processes*. Wiley, New York, pp. 167–180
- Heath, C., Tversky, A. (1991) : « Preference and Belief : Ambiguity and Competence in Choice under Uncertainty ». *Journal of Risk and Uncertainty*. Vol. 4, p 5-28.
- Heidegger, M. (1927) : *Être et temps*. Traduction française E. Martineau (1985). Édition numérique hors commerce.
- Heidegger, M. (1966) : *Temps et Être*. In *Questions III et IV*. Traduction française Jean Beaufret et Al. Tel Gallimard.
- Hermalin, B. E., Isen, A. M. (2006) : « A Model of the Effect of Affect on Economic Decision Making ». *Quantitative Marketing and Economics*. Vol. 6, Issue I, p 17-40.
- Herstein, I. N., Milnor, J. (1953) : « An axiomatic approach to measurable utility ». *Econometrica*. Vol. 21, 291–297.
- Hertwig, R., Gigerenzer, G. (1999) : « The 'Conjunction Fallacy' Revisited : How Intelligent Inferences Look Like Reasoning Errors ». *Journal of Behavioral Decision Making*. Vol. 12, p 275-305.
- Holton, R. (1999) : « Intention and Weakness of Will ». *Journal of Philosophy*. Vol. 96, p 241-262.
- Hsee, C. K. (1996) : « The Evaluability Hypothesis : An Explanation for Preference Reversal between Joint and Separate Evaluations of Alternatives ». *Organizational, Behavior and Human Decision Processes*. Vol. 67, No. 3, p 247-257.
- Hsee, C. K., Loewenstein, G. F., Blount, S., Bazerman, M. H. (1999) : « Preference Reversals between Joint and Separate Evaluation of Options : A Review and Theoretical Analysis, » *Psychological Bulletin*. Vol. 125, p 576–590.
- Huber, O., Kunz, U. (2006) : « Time Pressure in risky decision-making : effect on risk defusing ». *Psychology Science*. Vol. 49, p 415-426.
- Hume, (1739/1740) : *A treatise of human nature*. Dover Philosophical Classics, (2003).
- Husserl, E. (1905-1935) : *Sur l'intersubjectivité*. Traduction française Depraz, N. (2001). P.U.F, Épiméthée.
- Husserl, E. (1913) : *Idées directrices pour une phénoménologie et une philosophie phénoménologique pures*. Tome 1. Traduction française Ricoeur, P. (1950). Tel Gallimard.

- Husserl, E. (1952): *Idées directrices pour une phénoménologie et une philosophie phénoménologique pures*. Livre second. Traduction Française Escoubas, E. (1982). P.U.F, Épiméthée.
- Hutchinson, J. M. C., Gigerenzer, G. (2005) : « Simple Heuristics and Rules of Thumb : where Psychologists and Behavioural Biologists Might Meet ». *Behavioural Processes*. Vol. 69, p 97-124. Reprinted in Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011), p 110-133.
- Hwang, M. L. (1994) : « Decision making under time pressure : A model for information systems research ». *Information & Management*. Vol. 27, p 197-203.
- Jackson, F. (1984) : « Weakness of Will ». *Mind*. Vol. 93, p 1-18
- Jeffrey, R. C. (1976) : « Savage's omelet ». *Proceedings of the Biennial Meeting of the Philosophy of Science Association*. p 361-371.
- Jeffrey, R.C. (1965) : *The Logic of Decision*. Second Edition, 1983. University of Chicago Press.
- Jeffrey, R. C. (2004) : *Subjective Probability: The Real Thing*. Cambridge University Press.
- Johnson, E. J., Hershey, J., Meszaros, J., Kunreuther, H. (1993) : « Framing, Probability Distortions, and Insurance Decisions ». *Journal of Risk and Uncertainty*. Vol. 7, p 35-51.
- Kahneman, D., Tversky, A. (1972) : « Subjective probability : a judgment of representativeness ». *Cognitive Psychology*. Vol. 3, p 430-454. Reprinted in Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982), p 32-49.
- Kahneman, D., Tversky, A. (1973) : « On the psychology of prediction ». *Psychological Review*. Vol. 80, p 237-251. Reprinted in Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982), p 48- 68.
- Kahneman, D., Tversky, A. (1979) : « Prospect Theory. An Analysis of Decision under Risk ». *Econometrica*. Vol. 47, No. 2, p 263-291. Reprinted in Kahneman, D., Tversky, A. (Eds.) (2000), p 17-43.
- Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982) : *Judgment Under Uncertainty : Heuristics and Biases*. Cambridge University Press.
- Kahneman, D., Tversky, A. (1984) : « Choices, Values, and Frames ». *American Psychologist*. Vol. 39, No. 4, p 341-350. Reprinted in Kahneman, D., Tversky, A. (Eds.) (2000), p 1-16.
- Kahneman, D., Knetsch, J. L., Thaler, R. H. (1991) : « The Endowment Effect, Loss Aversion, and Status Quo Bias ». *Journal of Economic Perspectives*. Vol. 5, No. 1, p 193-206. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 159-179.
- Kahneman, D. (1993) : « New Challenges to the Rationality Assumption ». *Journal of Institutional and Theoretical Economics*. Vol. 150, No. 1, p 18-36. In Kahneman, D., Tversky, A. (Eds.) (2000), p 758-774.
- Kahneman, D. (2000) : « Experienced Utility and Objective Happiness ». In Kahneman, D., Tversky, A. (Eds.) (2000), p 673-692.
- Kahneman, D. (2000b) : « Evaluation by Moments. Past and Future ». In Kahneman, D., Tversky, A. (Eds.) (2000), p 693-708.
- Kahneman, D., Tversky, A. (Eds.) (2000) : *Choices, Values, and Frames*. Cambridge University Press.
- Kahneman, D. Frederick, S. (2002) : « Representativeness Revisited : Attribute Substitution in Intuitive Judgment ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 49-81.
- Kahneman, D. (2010) : *Thinking Fast and Slow*. Anchor Canada.

- Kant, I. (1781/1787) : *Kritik der reinen Vernunft*. Felix Meiner Verlag, (1998).
- Karni, E. (2009) : « Savage's Subjective Expected Utility Model ». In Durlauf, S. N., and Blume, L. E. (2008) : *The New Palgrave Dictionary of Economics*, Second Edition.
- Katsikopoulos, K. V., Gigerenzer, G. (2008) : « One-reason decision-making: Modeling violations of expected utility theory ». Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011), p 186-200.
- Kauder, E. (1965) : *A History of marginal utility*. Princeton University Press.
- Keren, G., Teigen, K. H. (2004) : « Yet Another Look at the Heuristics and Biases Approach ». In Koehler, D. J, Harvey N. (Eds.) (2004), p 89-110
- Keynes, J. M. (1921) : *A Treatise on Probability*. Macmillan (1963)
- Kimball, M. S. (1993) : « Standard Risk Aversion ». *Econometrica*. Vol. 61, No. 3, p 589-611.
- Knight, F. (1963) : *Risk, Uncertainty, and Profit*. Houghton Mifflin Company.
- Koehler, D. J., Harvey, N. H. (eds.) (2004) : *Blackwell Handbook of Judgment & Decision Making*. Blackwell Publishing.
- Koopman, B. O. (1940) : « The Axioms and Algebra of Intuitive Probability ». *The Annals of Mathematics*. 2nd Ser., Vol. 41, No. 2, p 269-299.
- Kraft, C. H., Pratt J. W., Seidenberg, A. (1959) : « Intuitive Probability on finite Sets ». *Ann. Math. Stat.* Vol. 30, No. 2, p 408-419.
- Kranz, D. H., Luce, R. D., Suppes, P., Tversky, A. (1971) : *Foundations of Measurement 3* vol. Dover Books of Mathematics.
- Kreps, D. M. (1988) : *Notes on the Theory of Choice*. Westview Press, Boulder and London.
- Kripke, S. (1962): « The Undecidability of Monadic Modal Quantification Theory ». *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*. Vol. 8, p 113–116.
- Kripke, S. (1963) : « Semantical Considerations on Modal Logic ». *Acta Philosophica Fennica*. Vol.16, p 83–94.
- Kuhl, J. (1987) : « Action control: The maintenance of motivational states ». In Halisch, F., Kuhl, J. (Eds.) (1987) : *Motivation, Intention and Volition*, p 279-291. Berlin: Springer.
- Kyburg, H. E., Smokler, H. E. (1980) : *Studies in subjective probability*. Huntington. N.Y.
- Lafortune, J-M. (2004) : *Introduction aux analyses sociologiques du temps hors travail. Fondements théoriques et enjeux sociaux du temps libre, du loisir, du jeu et du sport*. Québec : PUQ.
- Largeault, J. (1979) : *Hasards, probabilités, inductions : petits écrits de circonstance*. Association des publications de l'Université de Toulouse-Le Mirail.
- Larrère, C. (1992) : *L'invention de l'économie au XVIIIe siècle*. P.U.F, Léviathan.
- Leibniz, G. W. : *Sämtliche Schriften und Briefe*, Ed. Deutsche Akademie der Wissenschaften, Berlin (1923 -).
- Lepage, F., Morgan, C. (2011) : « Revision with Conditional Probability Functions: Two Impossibility Results » in P. Girard et Al. (eds.), *Dynamic Formal Epistemology, Synthese Library*. Vol. 35, p 161-172.
- Lévinas, E. (1971) : *totalité et infini*. Livre de Poche (2009).
- Levy, J. S. (1992) : « An Introduction to Prospect Theory ». *Political Psychology*. Vol. 13, No. 2, p 171-186.
- Levy, J. S. (2003) : « Application of Prospect Theory to Political Science ». *Synthese*. Vol. 135, p 215-241.
- Lewis, D. (1983) : *Philosophical Papers*. Vol. 1. Oxford University Press.

- Lewis, D. (1986) : *On the plurality of worlds*, Oxford, Blackwell.
- Lichtenstein, S., Slovic, P. (1971) : « Reversals of Preference Between Bids and Choices in Gambling Decisions ». *Journal of Experimental Psychology*. Vol. 89, p 46-55. Reprinted in Lichtenstein, S., Slovic, P. (eds.) (2006), p 69-76.
- Lichtenstein, S., Slovic, P. (1973) : « Response-Induced Reversals of Preferences in Gambling : an Extended Replication in Las Vegas. *Journal of Experimental Psychology*. Vol. 101, p 16-20. Reprinted in Lichtenstein, S., Slovic, P. (eds.) (2006), p 52-68.
- Lichtenstein, S., Slovic, P. (Eds.) (2006) : *The Construction of Preference*. Cambridge University Press.
- Loewenstein, G. F., Prelec, D. (1992) : « Anomalies in Intertemporal Choice ». *The Quarterly Journal of Economics*. Vol. 107, No. 2, p 573-597. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 578-596.
- Loewenstein, G. F., Prelec, D. (1993) : « Preferences for Sequences of Outcomes ». *Psychological Review*. Vol. 100, No. 1, p 91-108. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 565-577.
- Loewenstein, G., Lerner, J. S. (2003) : « The role of affect in decision making ». In Davidson, R. J., Scherer, K. R., Goldsmith, H. H. (2003) : *Handbook of Affective Science*. Oxford University Press.
- Long, A. A., Sedley, D. N. (1987) : *Les philosophes hellénistiques*. Traduction française Brunschwig, J., Pellegrin, P. (2001). GF Flammarion.
- Loomes, G., Sugden, R. (1982) : « Regret Theory : An Alternative Theory of rational Choice under Uncertainty ». *The Economic Journal*. Vol. 92, p 805-824.
- Luce, R. D., Raiffa, H. (1957) : *Games and Decisions*. Dover.
- Luce, M. F., Bettman J. R., Payne, J. W. (1997) : « Choice Processing in Emotionally Difficult Decisions ». *Journal of Experimental Psychology*. Vol. 23, No. 2, p 384-405.
- Luce, M. F., Payne, J. W., Bettman J. R. (1999) : « Emotional Trade-Off Difficulty and Choice ». *Journal of Marketing Research*. Vol. 36, No. 2, p 143-159.
- Machina, M. J. (1982) : « ‘Expected utility’ analysis without the independence axiom ». *Econometrica*. Vol. 50, No 2, p 277-323.
- Machina, M. J. (1987) : « Choice Under Uncertainty : Problems Solved and Unsolved ». *Economic Perspective*. Vol. 1, No. 1, p 121-151.
- Malinvaud, E. (1952) : « Note on von Neumann-Morgenstern’s Strong Independence Axiom ». *Econometrica*. Vol. 20, Issue 4, p 679.
- Manktelow, K. (2012) : *Thinking and Reasoning*. Psychology Press. London and New York.
- Manzini, P., Mariotti, M. (2009) : « Choice over time ». in Anand, P., Pattanaik, P. K., Puppe, C. (Eds.) (2009), p 239-271.
- Markovitz, H. (1952) : « The Utility of Wealth ». *Cowles Comission Paper*. New series, No. 57, p 151-158.
- Marr, D. (1982) : « Representation and recognition of the movements of shapes ». *Proceedings of the Royal Society of London*. B 454, p 501-524.
- Marschak, J. (1950) : « Rational Behavior, Uncertainty Prospects, and Measurable Utility ». *Econometrica*. Vol. 18, No. 2, p 111-141.
- Martin, T. (Dir.) (2003) : *Probabilités subjectives et rationalité de l’action*. CNRS : Paris.
- Maule, J.A., Hockey, G.R., Bdzola, L. (2000) : « Effects of time-pressure on decision making under uncertainty : changes in affective state and information processing strategy ». *Acta Psychologica*. Vol. 104, p 283-301.

- May, J., Holton R. (2010) : « What in the World is Weakness of Will ? ». *Philo Stud.* DOI 10.1007/s 11098-010-9651-8. Springer online Publishing.
- Maynard Smith, J. (1979) : « Game theory and the evolution of behaviour ». *Proc. Roy. Soc. Lon.* B 205, p 475-488.
- McClennen, E. F. (1983) : « Sure-thing doubts ». *Foundations of Utility and Risk Theory with Applications*. Reprinted in Gärdenfors, P., Sahlin, N-E. (Eds.) (1988), p 166-182.
- McClennen, E. F. (1990) : *Rationality and Dynamic choice*. Cambridge University Press.
- McDermott, R., Fowler, J. H., Smirnov, O. (2008) : « On The Evolutionary Origin of Prospect Theory Preferences ». *The Journal of Politics*. Vol. 70, No. 2, p 335-350.
- Mele, A. (2009) : « Weakness of Will and Akrasia ». *Philo Stud.* DOI 10.1007/s 11098-009-9418-2. Springer online Publishing.
- Merleau-Ponty, M. (1945) : *Phénoménologie de la perception*. Tel, Gallimard.
- Mongin, P. (2011) : « La théorie de la décision et la psychologie du sens commun ». *Cahier de recherche GREGHEC*. No 943.
- Morgenstern, O. (1979) : « Some Reflections on Utility ». In Allais, M., Hagen, O. (Eds) (1979), p 175-183.
- Mosteller, F., Nogee, P. (1951) : « An experimental measurement of utility ». *Political Economy*. Vol. 59, p 371-404.
- Myerson, R. B. (1991) : *Game Theory : analysis of conflict*. Harvard University Press.
- Nash, J.F. (1950) : «Equilibrium Points in n -Person Games ». *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 36, No. 1, p 48-49.
- Nietzsche, F. (1887) : *Généalogie de la morale*. Traduction française, Blondel, É. et Al. (1996). Flammarion, Mille & une pages, (2000)
- Nozick, R. (1969) : « Newcomb's problem and two principles of choice ». In *Essays in Honor of Carl G. Hempel*, 1969.
- Nygren, T. E., Isen, A. M., Taylor, P. J., Dulin, J. (1996) : « The Influence of Positive Affect on the Decision Rule in Risk Situations : Focus on Outcome (And Especially Avoidance of Loss) Rather than Probability ». *Organizational Behavior and Human Decision Processes*. Vol. 66, No. 1, p59-72.
- Oaksford, M., Chater, N. (2007) : *Bayesian rationality*. Oxford University Press.
- Ore, O. (1953) : *Cardano. The Gambling Scholar*. Dover.
- Paquette, M. (2006) : *La délibération et les théories axiomatisées de la décision*. Thèse présentée à l'Université du Québec à Trois-Rivières.
- Pascal, B. *Œuvres Complètes*. Le Seuil, 1963.
- Payne Bettman adaptative.
- Pearson, E. S., and M G Kendall, M. G. (Eds.) (1955) : *Studies in the history of Probability and statistics*. Vol. 1. *Biometrika*.
- Peterson, M. (2009) : *An Introduction to Decision Theory* Cambridge University Press.
- Picavet, E. (1996) : *Choix rationnel et vie publique*. P.U.F, Essais.
- Poundstone, W. (1992) : *Prisoner's Dilemma*. Anchor Books, New York.
- Platon, *Protagoras*. Traduction française Robin, L. (1950), Bibliothèque de la Pléiade.
- Platon, *Timée*. Traduction française Robin, L. (1950), Bibliothèque de la Pléiade.
- Pratt, J. W. (1964) : « Risk Aversion in the Small and in the Large ». *Econometrica*. Vol. 32, No. 1/2 , p 122-136.
- Prelec, D. (2000) : « Compound Invariant Weighting Functions in Prospect Theory ». In Kahneman, D., Tversky, A. (Eds.) (2000), p 67-92.

- Proust, J. (2005) : *La nature de la volonté*. Gallimard, Folio Essais.
- Proust, J. (2007) : « Metacognition and metarepresentation : Is a self-directed theory of mind a precondition for metacognition? ». *Synthese*.
- Proust, J. (2010) : *Les animaux pensent-ils ?* Bayard.
- Quiggin, J. (1982) : « A theory of anticipated utility ». *Journal of Economic Behavior and Organization*. Vol. 3, p 323-343.
- Rabin, M., Thaler, R. (2001) : « Anomalies : Risk Aversion ». *The Journal of Economic Perspectives*. Vol. 15, No. 1, p 219-232.
- Raiffa, H. (1968) : *Decision analysis; introductory lectures on choices under uncertainty*. Don Mills, Ont., Addison-Wesley.
- Ramsey, F. P. (1926) : « Vérité et probabilité ». In, Engel, P., Marion, M. (Eds) : *Logique, philosophie et probabilités*. Vrin, Mathesis, 2003.
- Read, D. (2004) : « Intertemporal Choice ». In Koehler, D. J., Harvey, N. H. (Eds.) (2004), p 424-443.
- Ricoeur, P. (1950) : *Philosophie de la volonté. Tome I. Le Volontaire et l'Involontaire*. Points, Essais. Réédition (2009).
- Ricoeur, P. (1986) : *Du texte à l'action*. Points, Essais.
- Ricoeur, P. (1990) : *Soi-même comme un autre*. Points, Essais.
- Romano, C. (1999) : *L'Événement et le Temps* Paris, PUF, « Épiméthée »
- Romano, C. (2010) : *Au cœur de la raison : la phénoménologie*. Folio, Essais.
- Ryan, R. M., Deci, L. D. (2000) : « Intrinsic and Extrinsic Motivations : Classic Definitions and new Directions ». *Contemporary Educational Psychology*. Vol. 25, p 54-67.
- Ryan, R. M., & Deci, E. L. (2002). An overview of self-determination theory. In Ryan, R. M., Deci, E. L. (2002) (Eds.), *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.
- Sahlin, N-E. (1990) : *The philosophy of F. P. Ramsey*. Cambridge University Press.
- Saint-Sernin, B. (1973) : *Les mathématiques de la décision*. P.U.F.
- Samuelson, P. A. (1952) : « Probability, Utility, and the Independence Axiom ». *Econometrica*. Vol. 20, No. 4, p 670-678.
- Savage, L. J. (1951) : « The theory of statistical decision ». *Journal of the American Statistical Association*. Vol. 46, p 55-67
- Savage, L. J. (1954) : *The Foundations of Statistics*. Second revised Edition (1972). Dover.
- Savage, L. J. (1960) : « The foundations of Statistics Reconsidered ». *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. p 575-586.
- Savage, L. J. (1967) : « Difficultés de la Théorie de la Probabilité Personnelle ». (Traduction, Morlat, G.). *Mathématiques et Sciences Humaines*. Tome 21, p 5-9.
- Savage, L. J. (1968) : « Probabilité Personnelle et Induction ». (Trad, Vaguey, G. et Morlat, G.). *Mathématiques et Sciences Humaines*. Tome 23, p 5-15.
- Schmeidler, D. (1989) : « Subjective probability and expected utility without additivity ». *Econometrica*. Vol. 57, p 571-587.
- Schmeltzer, C., Caverni, J-P., Warglien, M. (2008) : « Le renversement des préférences...renversé ! ». *Psychologie Française*. Vol. 53, p 39-50.
- Schwarz, N. (2002) : « Feeling as Information : Moods Influence Judgments and Processing Strategies ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 534-547.
- Shafer, G. (1986) : « Savage Revisited ». *Statistical Science*. Vol. 1, No. 4, p 463-501.
- Simon, H. A. (1955) : « A Behavioral Model of Rational Choice ». *The Quarterly Journal of Economics*. Vol. 59, No. 1, p 99-118.

- Simon, H. A. (1956) : « Rational Choice and the Structure of Environment ». *Psychological Review*. Vol. 63, p 129-138.
- Simon, H. A. (1959) : « Theories of Decision-Making in Economics and Behavioral Science ». *The American Economic Review*. Vol. 49, No. 3, p 253-283.
- Simon, H. A. (1991) : « Bounded Rationality and Organizational Learning ». *Organization Science*. Vol 2, Issue 1, p125-134.
- Skyrms, B. (1986) : *Choice & Chance*. Wadsworth Publishing Company.
- Sloman, S. A. (2002) : « Two Systems of Reasoning ». In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 379-396.
- Slovic, P., Lichtenstein, S. (1968) : « Relative Importance of Probabilities and Payoffs in Risk Taking ». *Experimental Psychology Monograph*. Vol. 78. No. 3, p 1-18. Reprinted in Lichtenstein, S., Slovic, P. (eds.) (2006), p 41-51.
- Slovic, P., Lichtenstein, S. (1983) : « Preference Reversals : A broader Perspective ». *The American Economic Review*. Vol. 73, No. 4, p 596-605.
- Slovic, P. (1991) : « The Construction of Preference ». *American Psychologist*. Vol. 50, No. 5, p 364-371. Reprint in Kahneman, D., Tversky, A. (eds.) (2000), p 489-502.
- Slovic, P., Peters, E., Finucane, M. L., MacGregor, D. J. (2005) : « Affect, Risk, and Decision Making ». *Health Psychology*. Vol. 24, No. 4, p S35-S40.
- Spinoza, B. (1677) : *Éthique*. Traduction française, Appuhn, C. (1929). GF Flammarion.
- Steel, P. (2007) : « The Nature of Procrastination : A meta-Analytic Review of Quintessential Self-Regulatory Failure ». *Psychological Bulletin*. Vol. 133, No. 1, p 65-94.
- Stich, S. (1983) : *From Folk Psychology to Cognitive Science. The Case against Belief*, Cambridge, Ma., Bradford Books & The MIT Press.
- Stickney, L. T. (2009) : « Affect and Decision making ». *Decision Line*, p 4-6.
- Stigler, G. L. (1950) : « The Development of Utility Theory. I ». *The Journal of Political Economy*. Vol 58, No 4, p 307-317.
- Stigler, G. L. (1950) : « The Development of Utility Theory. II ». *The Journal of Political Economy*. Vol. 58, No. 5, p 373-396.
- Stroud, S., Tappolet, C. (Eds.) (2003) : *Weakness of Will and Practical Irrationality*. Oxford University Press.
- Sugden, R. (1991) : « Rational Choice: A Survey of Contributions from Economics and Philosophy ». *Royal Economic Society, The Economic Journal*. Vol. 101, Issue 407, p 751-785.
- Sugden, R. (2004) : « Alternatives to Expected Utility : Foundations ». In Barberà, S., Hammond, P. J., Seidl, C. (Eds.) (2004), p 685-756.
- Suppes, P., Zinnes, J. L. (1963) : *Basic measurement theory*. TECHNICAL REPORT NO. 45 March 15, 1962. Institute for mathematical studies in the social sciences, Stanford University.
- Suppes, P. (1972) : *Axiomatic Set Theory*. Dover.
- Suppes (1994) : « Qualitative Theory of Subjective Probability ». In Wright, G., Ayton, P. (1994) : *Subjective Probability*. John Wiley & Sons Ltd.
- Surowik, D. (2002) : « Leonard Savage's Mathematical Decision Theory ». *Studies in Logic, Grammar and Rhetoric*. Vol. 5, No. 18, p 65-75.
- Tappolet, C. (1995) : « Les émotions et les concepts axiologiques ». In Ogien, R., Paperman, P. (Dir.) (1995) : *La couleur des pensées*. p 237-257.
- Tappolet, C. (2000) : *Émotions et valeurs*. P.U.F.
- Tappolet, C. (2003) : « Emotions and the Intelligibility of Akratic Action ». In Stroud, S., Tappolet, C. (Eds.) (2003), p 97-120.

- Thaler, R. H. (1980) : « Toward a Positive Theory of Consumer Choice ». *Journal of Economic Behavior and Organization*. Vol. 1, p 39-60. Reprint in Kahneman, D., Tversky, A. (eds.) (2000), p 269-287.
- Thaler, R. H. (1999) : « Mental Accounting Matters ». *Journal of Behavioral Decision Making*. Vol. 12, p 183-206. Reprinted in Kahneman, D., Tversky, A. (eds.) (2000), p 241-268.
- Tiles, M. (1989) : *The Philosophy of set theory*. Dover.
- Tversky, A., Kahneman, D. (1971) : « Belief in the law of small numbers ». *Psychological Bulletin*. Vol. 2, p 105-110. Reprinted in Kahneman, D., Slovic, P., Tversky, A. (eds.) (1982), p 23-31.
- Tversky, A. (1972) : « Elimination by aspects: A theory of choice ». *Psychological Review*. Vol. 79, No. 4, p 281-299.
- Tversky, A., Kahneman, D. (1973) : « Availability : a heuristic for judging frequency and probability ». *Cognitive Psychology*. Vol. 4, p 207-232. Reprinted in Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982), p 164-178.
- Tversky, A., Kahneman, D. (1974) : « Judgment under uncertainty : heuristics and biases ». *Science*. Vol. 185, p 1124-1131. Reprinted in Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982), p 3-20.
- Tversky, A., Kahneman, D. (1982) : « Evidential impact of base rate ». In Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982), p 153-160.
- Tversky, A., Kahneman, D. (1984) : « Extensional versus Intuitive Reasoning : The Conjunction Fallacy in Probability Judgment. *Psychological Review*. Vol. 91, p 293-315. Reprinted in Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 19-48.
- Tversky, A., Kahneman, D. (1986) : « Rational Choice and the Framing of Decisions ». *Journal of Business*. Vol. 59, No. 4, p 5251-5278. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 209-223.
- Tversky, A., Kahneman, D. (1991) : « Loss Aversion in Riskless Choice ». *The Quarterly Journal of Economics*. Vol. 106, No. 4, p 1039-1061. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 143-158.
- Tversky, A., Kahneman, D. (1992) : « Cumulative Prospect Theory : An Analysis of Decision under Uncertainty ». *Journal of Risk and Uncertainty*. Vol. 5, p 297-323. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 44-65.
- Tversky, A., Koehler, D. J. (1994) : « Support Theory : a Nonextensional Representation of Subjective Probability ». *Psychological Review*. Vol. 101, p 547-567. Reprinted in Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002), p 441-473.
- Tversky, A., Fox, C. R. (1995) : « Weighting Risk and Uncertainty ». *Psychological Review*. Vol. 102, No. 2, p 269-283. Reprint in Kahneman, D., Tversky, A. (Eds.) (2000), p 93-117.
- Tversky, A., Wakker, P. (1995) : « Risk Attitudes and Decision Weights ». *Econometrica*. Vol. 63, No. 6, p 1255-1280.
- Vickers, J. M. (1962) : *A critical investigation of Frank Ramsey's theory of value and belief*. Ph. D., Thesis, Stanford University.
- Vickers, J. M. (1988) : *Chance and Structure: An Essay on the Logical Foundations of Probability*. Oxford University Press.
- Vickrey, W. (1945) : « Measuring Marginal Utility by Reaction to Risk ». *Econometrica*. Vol. 13, Issue 4, p 319-333.
- von Neumann, J., Morgenstern, O. (1947): *Theory of games and economic behavior*. Princeton University Press, Princeton, NJ.

- Wakker, P. P. (2010) : *Prospect Theory*. Cambridge University Press.
- Weiner, B. (2005) : « Motivation from an Attribution Perspective and the Social Psychology of Perceived competence ». In Elliot, A. J., Dweck, C. S. (Eds.) (2005), p 73-84.
- Weirich, P. (2004) : « Realistic Decision Theory ». Oxford University Press.
- Weiss, J. W., Weiss, D. J. (eds.) (2009) : *A Science of Decision making. The Legacy of Ward Edwards*. Oxford University Press.
- Willemsen, M.C., Keren, G. (2004) : « The role of negative features in joint and separate evaluation ». *J. Behav. Dec. Making*. Vol. 17, p 313–329.
- Williams, A. C. (1966) : « Attitudes toward speculative risks as an indicator of attitudes toward pure risks ». *Journal of Risk and Insurance*. Vol. 33, p 577-586.
- Willinger, M. (1990) : « La rénovation des fondements de l'utilité et du risque ». *Revue Économique*. Vol. 41, No.1, p 5-48.
- Wittman, M., Paulus, M. P. (2007) : « Decision making, impulsivity and time perception ». *TRENDS in Cognitive Sciences*. Vol. 30, No. 10, p 1-6.
- Wu, G., Zhang, J., Gonzalez, R. (2004) : « Decision Under Risk ». In Koehler, D. J., Harvey, N. H. (Eds.) (2004), p 399-423.
- Yaqub, M. Z., Saz, G., Hussain, D. (2009) : « A Meta Analysis of the Empirical Evidence on Expected Utility Theory ». *European Journal of Economics, Finance and Administrative Science*. Issue 15, p 117-133.
- Yuen, K. S. L, Lee, T. M. C. (2002) : « Could mood state affect risk-taking decision ? ». *Journal of Affective Disorders*. Vol. 75, p 11-18.
- Zajonc, R. B. (1980) : « Feeling and Thinking. *Preferences Need No Inferences* ». *American Psychologist*. Vol. 35, No. 2, p 151-175.